APR 25 2006

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	20.Apr.06	DISSERTATION

4. TITLE AND SUBTITLE
GEOGRAPHIC DIFFERENTIATION OF FRANCISELLA TULARENSIS USING MOLECULAR METHODS.

5. FUNDING NUMBERS

6. AUTHOR(S)
MAJ DEMPSEY MICHAEL P

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
UNIVERSITY OF NEBRASKA MEDICAL CENTER

8. PERFORMING ORGANIZATION REPORT NUMBER
CI04-1767

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
THE DEPARTMENT OF THE AIR FORCE
AFIT/CIA, BLDG 125
2950 P STREET
WPAFB OH 45433

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Unlimited distribution
In Accordance With AFI 35-205/AFIT Sup 1

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

14. SUBJECT TERMS

15. NUMBER OF PAGES
192

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

Geographic Differentiation of *Francisella tularensis* Using Molecular Methods

Michael Patrick Dempsey, Ph.D.

University of Nebraska, 2006

Advisor: Steven H. Hinrichs, M.D.

Francisella tularensis is a category-A bio-threat agent posing significant concern to both warfighter and civilian. The four subspecies are t*ularensis*, *holarctica*, *mediaasiatica*, and *novicida* with *tularensis* being most virulent in humans. *F. tularensis* species have been isolated throughout the Northern Hemisphere, but subsp. *tularensis* isolates appear restricted to North America. The research described here involved identification and characterization of genetic differences between *F. tularensis* subspecies and the correlation of genetic markers with geographic variation. The *F. tularensis* genome is highly conserved among all subspecies when evaluated by such methods as pulsed-field gel electrophoresis, but newer molecular methods offer potential for higher resolution. The present work was based on previous comparative genomic hybridization (CGH) studies that identified several regions of difference (RD) between *F. tularensis* subspecies. The working hypothesis was that transposon-associated insertion-sequence (IS) elements are the primary driving factor in *F. tularensis* subspecies divergence. Analysis of these RD showed that only a small number of genes within subsp. *tularensis* are absent from subsp. *holarctica*, and all RD were associated with IS elements. The second hypothesis was that geographic-specific subpopulations can be differentiated with advanced molecular methodologies. CGH-testing of a large global *F. tularensis* strain collection resulted in discovery of a novel IS-associated RD within *holarctica* strains, referred to as RD_{Spain}. Confirmation of this finding was demonstrated by PCR analysis of a global DNA repository, and RD_{Spain} was found to be restricted to Spain and France. Paired-end sequence mapping (PESM) was used to catalogue additional candidate differential genes. PESM revealed 17 contiguous regions (CR) within *holarctica* ($CR_{holarctica}$) having extensive IS-mediated genome rearrangements within

corresponding *tularensis*-specific sequence regions. Several CR demonstrated altered genes potentially explaining some subspecies-specific virulence and biochemical differences. Nested-PCR testing demonstrated CR-conservation in spatially and temporally diverse strains of each subspecies. PESM also identified additional geographic-specific subtypes including two isolates potentially representing a new *F. tularensis* taxonomic unit. These studies demonstrated that IS-element driven mechanisms were responsible for subspecies divergence and provided models to improve understanding of molecular and geographic divergence. Further, the studies culminated in a novel PCR subspecies-subtyping strategy for application to field work.

GEOGRAPHIC DIFFERENTIATION OF *FRANCISELLA TULARENSIS* USING MOLECULAR METHODS

By

Michael Patrick Dempsey

B.A., University of Nebraska at Omaha, 1985
MT (ASCP), University of Nebraska Medical Center, 1992
M.S., University of Nebraska Medical Center, 1993

A DISSERTATION

Presented to the Faculty of

The Graduate College in the University of Nebraska

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

Medical Sciences Interdepartmental Area
Pathology and Microbiology

Under the Supervision of Professor Steven H. Hinrichs, MD

Medical Center
Omaha, Nebraska

May, 2006

APPROVAL PAGE

ACKNOWLEDGEMENTS

I am extremely grateful to the United States Air Force for funding me for the last three years – who, while at my age and with a family, afforded me an opportunity I otherwise would not have had toward completing this degree. It is necessary to point out that the views expressed in this dissertation are those of the author, and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U. S. Government.

In addition, I wish to thank my many collaborators who helped in the completion of this work. Besides all the excellent staff at the AFIP, I am especially grateful to Dr. Michael Dobson, John McGraw, Mark Chrustowski, and Jennifer Engle for their assistance in the Genomics laboratory. I also thank Dr. Steve Francesconi, Bob Burgess, and Wendell Thomas, as well as Bob Wickert from UNMC, for their assistance with *Francisella* strain cultivation and DNA panel preparations. Also from UNMC, I thank Michael Olsen for doing nearly all the "MLVA" work. From UNL, I thank Joe Nietfeldt for "PESM" library construction, Chao and Min Zhang for microarray help, and John Wise for administering Pathogene-server access. From Northern Airzona University, I am grateful to Dr. Paul Keim, Miles Stanley, Julia Rhodes, Dr. Dave Wagner, and Matt Van Ert for their help with "RD_{Spain}" characterization. From The Institute of Genomics Research (TIGR), I am very grateful to Jacques Ravel for his bioniformatics and illustration expertise. I am also grateful to Dr. Emilio Garcia, from Los Alamos National Laboratories, for use of his *F. tularensis* "LVS" sequence. I also thank Col Steve Putbrese who collaborated with Dr. Bernd Jilly from the Alaskan State Public Health Laboratory to contribute three *F. tularensis* strains to our collection.

From the administrative front-line, Levi Horton and Mary Ann King have heroically processed numerous supply orders and managed my countless administrative details. Also, I am appreciative to Dr. Matt Sharkey for his editorial review of chapters and manuscripts.

In addition, I want to acknowledge my international collaborators. I thank Dr. César Gutiérrez Martin, from the Department of Animal Health, Facultad de Veterinaria, Leon, Spain,

for contribution of his entire *F. tularensis* strain DNA collection. I thank Dr. Christine Lion, from the Laboratoire de Bacteriologie, Centre Hospitalier et Universitaire, Nancy, France, for contributing two *F. tularensis* RD_{Spain} strains. For his extensive global *F. tularensis* DNA collection, I am grateful to Dr. Anders Johansson from the Division of Infectious Diseases and Clinical Bacteriology Department of NBC Analysis Defense Research Agency, Umea Sweden.

My academic advisors have been my mentors and provided unmatched encouragement and resourcefulness in helping me complete this degree in the Air Force's allotted time-frame. Although not one of my current advisors, I must thank Dr. Sam Cohen for his lasting encouragement and friendship of nearly twenty years now. He was instrumental in my recruitment back to Nebraska to initiate this course of study. Next I thank Dr. Steve Hinrichs who was on my Masters' graduate advisory committee, and who eagerly spear-headed to be my current advisory committee chairman. His creativity, patience, and encouragement, as well as availability at home on weekends, have greatly helped me stay on track. Next I thank Dr. Pete Iwen for his enthusiastic support and encouragement – and for asking me many tough questions to help build my "gray matter" during exams. I am indebted to Dr. Paul Fey's enthusiastic teaching of "Microbiology 898" where I was forever ruined from thinking that bacteria just behave themselves and never change – I suspect I will now always be skeptical as to the true genotype of any bacterium. I am extremely grateful for Dr. Andy Benson's arsenal of project ideas, and for his encouragement and confidence in my work, as well as his willingness to go for long runs in both Lincoln and Washington, DC. I am grateful to Dr. Debra Niemeyer who has been inspirational to me as a long-time mentor in the Air Force, and who helped stir my desire for advanced education – it's amazing what a couple of trips to the desert will do for one's creativity. I am extremely grateful to Dr. Bob Crawford for welcoming me back to AFIP, and for granting me an entire section of his laboratory "kingdom", including a never-ending supply of equipment and reagents for my own research. His kindness and inclusion of my family as part of his own (weddings and all) have been most appreciated.

I am grateful to my family and friends in Omaha for their care in provision of much encouragement and necessary, but enjoyable, study breaks during my year of study there. I am grateful to my parents for teaching me the value of persistence, hard work, and academic excellence. My parents-in-law were most kind by allowing me to live with them during the entire school year. For their sacrifice, I no longer consider them "in-laws", and I now look forward to staying with them longer and more often...and I'm sure they joyfully reciprocate my sentiments.

Now, I honor my wife and children. Children, I love you both and thank you for your sacrifice while I was away studying, and for how you helped your mother hold down the "fort". "Bubba-head", it has been odd having both of us in college at the same time – and I am certainly proud of how well you have done so far...but I would encourage you (and "Peanut-head") to finish all your education while you are young, and not as an old fart like me. Instead of studying or writing, I will certainly enjoy doing other activities with you including sporting events, watching movies (i.e., a Lord of the Rings Trilogy marathon) or playing video games. Now for my dear, sweet Wife – I love you! You are my source of sunshine and most treasured prize this side of Heaven. Thank you for your innumerable sacrifices, including managing the children and home during periods of intense study and late hours in the lab, and even during times of separation for months-on-end. I am extremely grateful for your editorial and formatting help with my dissertation. Hopefully now we can enjoy some time together without the distractions of studying or dissertation-editing.

Last, but to whom all glory is due, I thank God – my source of provision, creativity, and perseverance from beginning to end of this endeavor. He created me, redeemed me through His Son, Jesus, and has allowed me to enjoy many awesome discoveries of this world He created, some which are presented in this dissertation. I am also grateful for Church family in Maryland and Omaha who so lovingly cared for my family, especially during lengthy periods of separation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1:

Literature Review

1.0 Overview:

The objective of this chapter is to provide a comprehensive introduction to the bacterium *Francisella tularensis,* as well as the disease it produces -- tularemia, otherwise known as "rabbit fever" or "deer-fly fever". The first sections of this chapter will provide a classical background and history of *F. tularensis* including current knowledge surrounding its taxonomy and classification, ecology, virulence factors, and host-pathogen immune responses. The following concluding sections on molecular-genotyping and differentiation will provide an in-depth understanding of the molecular methodologies used thus far to differentiate between the multiple subspecies and geographical representatives of *F. tularensis*. The conclusion of this chapter will provide a launching point for the novel methodologies, results, and conclusions presented in this dissertation.

1.1 Background and History

1.1.1 Threat Significance

Francisella tularensis is a tiny, non-motile, faintly staining gram-negative coccobacillus originally isolated from ground squirrels in 1911 during a plague investigation in Tulare County, CA [1]. The organism is a facultative intracellular pathogen and is believed to affect more animal species, including humans, than any other known zoonotic pathogen [2, 3]. This organism has been weaponized and is considered a significant biowarfare agent, especially due to its ease of dissemination, its extremely low infectious dose of only ten to fifty organisms when acquired through the inhalation route in humans, and the potential existence of antibiotic-resistant strains that were genetically engineered under non-U.S. biological weapons programs [4-7].

Currently, *F. tularensis* is considered a Category-A Select Biological Agent of Human Disease [8]. In the wake of the 2001 anthrax attacks on the U.S. [9], the potential employment of *F. tularensis* as a weapon of bioterror, as well as its potential use as one of several biowarfare agents by the Iraqi military just prior to OPERATION Iraqi Freedom, has been strongly considered. In fact, as early as 1970 the World Health Organization (WHO) recognized the potential exploitation of *F. tularensis* to deliberately cause disease. WHO further predicted that illness would occur in as many as 50% of individuals receiving 25 or more bacterium from an attack using an antibiotic-sensitive strain, about half of those cases which would require hospitalization, and a 25% case-fatality rate would occur from such an attack [10, 11]. Due to this threat, rapid identification of *F. tularensis* following a potential covert release or during a naturally occurring outbreak is critical to both warfighter and civilian to 1) facilitate prompt action in limiting pathogen exposure, and 2) to ensure initiation of timely and specific post-exposure measures and treatments. The Centers for Disease Control (CDC) and Prevention has implemented Laboratory Response Network (LRN) diagnostic protocols [12], including culture, immunologic, and molecular methods for detecting *F. tularensis*, as have numerous other Government agencies including the DoD. Most of these existing methodologies, however, provide differentiation only to the species level; and as will be discussed later, differentiation to at least the subspecies level should be the goal of future identification strategies due to the clinical and forensic significance found at the subspecies and individual strain levels.

1.1.2 Taxonomy and Classification

The genus, *Francisella*, belongs to the γ-proteobacteria and is comprised of two species, *F. philomiragia* and *F. tularensis*. The species, *F. tularensis*, is comprised of four subspecies: subsp. *tularensis*, subsp. *holarctica*, subsp. *mediaasiatica*, and subsp. *novicida*. Subps. *mediaasiatica* and *novicida* exhibit moderate and low virulence, respectively, while only the highly virulent subsp. *tularensis* (Type A) and moderately virulent subsp. *holarctica* (Type B) are

clinically significant in humans [13, 14]. As reviewed by Chu and Weyant, at the species-level, both *F. philomiragia* and *F. tularensis* share morphological characteristics, have similar biochemical activities, and have a high degree of genetic relatedness [13]. Biochemically, both species are quite homogeneous but may be differentiated based on a few key tests (see table 1-1) such as *F. philomiragia* being oxidase-positive using Kovacs reagent unlike *F. tularensis* which is negative. Biochemical differentiation among *F. tularensis* subspecies may also be accomplished based on a few key tests including glycerol fermentation, glucose utilization, and citrulline ureidase activity. For example, subsps. *tularensis, novicida*, and *mediaasiatica* are similar in that they all utilize glycerol, and both subsps. *tularensis* and *mediaasiatica* have citrulline ureidase activity; but unlike subsps. *tularensis*, *holarctica*, and *novicida*, subsp. *mediaasiatica* is unable to utilize glucose. Subsp. *novicida* may be differentiated from the other subspecies by its ability to grow without cysteine supplementation and by its larger vegetative cell size (0.7-1.7 μM as compared to 0.2-0.7 μM for the other subsps.) [13]. Besides only minor biochemical differences, the high degree of genetic relatedness lends additional difficulty in explaining significant pathogenic differences among the *F. tularensis* subspecies [15]. With respect to pathogenesis, due to the extremely high risk of laboratory-acquired infection, especially involving subsp. *tularensis*, culturing of *F. tularensis* requires Biosafety Level-3 (BSL-3) containment [16] and is now often avoided. Currently, molecular-based methodologies allow for safe *F. tularensis* characterization from DNA preparations of killed bacteria, and as a result, many molecular-based research efforts including enhancing methods of subspecies differentiation, elucidation of molecular determinants of pathogenesis, and development of vaccine strategies are advancing. A detailed review of *F. tularensis* pathogenesis and molecular-characterization strategies will be presented later in this chapter.

1.1.3 Ecology of *F. tularensis*

F. tularensis is a widely infectious zoonotic pathogen and has been isolated from as many as 250 species of wildlife (reviewed by Oysten, Sjostedt, *et. al*) [10] including various lagamorphs, rodents, insectivores, carnivores, ungulates, marsupials, birds, amphibians, fish, and invertebrates (reviewed by Petersen and Schriefer) [17]. As reviewed by Chu and Weyent [13], habitats where lagamorphs (rabbits, hares, and Old World hares) and Rodentia (water voles, muskrats, lemmings, voles, and beavers) thrive are important in maintaining tularemia-enzootic foci; and biting arthropod vectors such as tabanid flies, ticks, and mosquitoes are considered important in mechanical tularemia transmission. Human acquisition occurs most often in association with hunting or other outdoor activities, by direct exposure with infected domesticated animals, or by bites from infected arthropod vectors. Recently, two disease cycles, terrestrial and aquatic (see table 1-2), have been described [18, 19]. As reviewed by Petersen and Schriefer [17], rabbits and hares often serve as amplifying hosts, and biting flies or ticks serve as arthropod vectors in the terrestrial cycle. For example in the United States, a correlation with exposures to animals, tick bites, and biting flies has been made for human disease in Western states (see table 1-3); but, whereas human cases in the central states have a similar correlation with the two former risk factors, such cases are rarely associated with biting flies [3, 18]. As for the aquatic cycle, beavers, muskrats, and voles serve as important mammalian hosts, and they appear to shed live organisms into their environments. Mosquitoes in Sweden have been strongly implicated as vectors in the transmission of from the aquatic cycle, but such a correlation has not been made for mosquitoes in the United States [17, 20]. Protozoa, such as *Acanthamoeba castellanii*, have recently been shown capable of harboring *F. tularensis*, and may play a significant role in maintenance of the organism in the aquatic cycle [21].

The geographic distribution of *F. tularensis* spans the entire Northern Hemisphere (see figure 1), with only a very recent isolated recovery of the organism from the Southern Hemisphere [22, 23]. Biochemical and molecular methods of subspecies differentiation have

shown that subsp. *tularensis* appears restricted to North America with the exception of rare isolates obtained from mites and fleas in Slovakia, Europe, but which haven't been associated with human cases [24]. Subsp. *holarctica* isolates occur both in North America (New World) as well as throughout the remainder of the N. Hemisphere (Old World). As reviewed by Petersen and Schriefer [17], subsp. *novicida* and subsp. *mediaasiatica* appear more focal in their distributions, with *novicida* exclusively isolated from North America except for the first case of its isolation from the S. Hemisphere (in Australia) as previously mentioned [23]. Subsp. *mediaasiatica* has been isolated only from the Central Asian (Kazakhstan and Turkmenistan) regions of the Former Soviet Union where it has been recovered from hares and ticks, but not humans [13, 25]. In addition, subsp. *holarctica* variant isolates exclusive to Japan (tentatively called subsp. *japonica*) [2, 25-27] and variant isolates apparently exclusive to Spain, France, and possibly Sweden, have been identified (Dempsey *et al.*, #1 in preparation), and will be discussed further in Chapter 3. Although numerous outbreaks have been reported worldwide, an outbreak in Spain between 1997 and 1998 provided an excellent collection of outbreak-isolate DNAs which have been the subject of several studies including a few described in this chapter, as well as in my own investigation presented in Chapter 3 of this dissertation. Certain epidemiological aspects of that particular outbreak are interesting and are presented in the next paragraph.

1.1.3.1 The 1997-1998 Spanish Outbreak:

As reviewed by Petersen and Schreifer [17], this was the first reported tularemia outbreak in Spain. In all, a total of 559 human cases of tularemia were reported, 519 of which came from Castille-Leon in Northwestern Spain. From a study of 142 patients from this region, 97.2% had indicated previous contact with hares; 83.3% had prepared hare carcasses, and 13.3% had handled hare meat [28]. Due to such high rates of cutaneous exposures, ulceroglandular tularemia was the most common form of clinical disease observed (87%); however, some cases of typhoidal, glandular, pneumonic, oculoglandular, and other atypical forms were also reported in humans

[28-31]. These clinical forms (see table 1-4 for a brief description of each form) of tularemia will be described in more detail in the next section. Also as expected, the isolates tested from affected humans, hares, ticks, and voles have been identified as subsp. *holarctica* [31-34]. It was also very interesting to note that of the 142 patients studied, 32 patients (22.5%) experienced intial treatment failure, most often associated with the ulceroglandular form of disease and use of doxycycline as the initial treatment. All 32 patients eventually responded favorably after a second round of treatment (and in a few cases, a third round was needed), the majority of whom responded to ciprofloxacin [28].

In 1998 a second human tularemia outbreak occurred in the central province of Cuenca, Spain, distant from the previous outbreak [35]. This time, nineteen cases of the ulceroglandular form were identified in individuals who had contact with crayfish. No isolates were recovered from this outbreak, and therefore no DNA was available for further testing (personal correspondence between P. Anda and this author); but positive Type-B 16S rDNA polymerase chain reaction (methodology is discussed in detail later in this chapter) results were obtained from the river, crayfish, and human lymph node aspirates indicating that those strains tested belong also to subsp. *holarctica*. Limited comparisons have demonstrated a possible minor difference in the outbreak-specific 16S rDNA sequences [33], but unlike the first outbreak, detailed phylogenetic analysis has not been possible for the second outbreak without isolates or DNA, and therefore the degree of genetic relationship between *F. tularensis* strains from the respective outbreaks, if any, has not been fully established.

1.1.4 Categories of Clinical Tularemia

Clinically, the onset of disease often occurs after an incubation period of approximately 3-6 days, and consists of symptoms often described as "flu-like" including fever, chills, malaise, headache, and sore throat; but more specific symptomology is variable and dependant on the route of entry. A consensus of the literature shows that there are six clinical forms of tularemia

[4, 7, 10, 17, 36], with the pneumonic form (accounting for only a rare percentage of cases, but which is of major concern from a biodefense perspective, especially involving subsp. *tularensis*,), being most severe with a case fatality rate as high as 30%-60% if untreated. The pneumonic form results from direct inhalation or from septicemic spread of infection from a non-pneumonic primary site. In recent years, this form of tularemia has been significantly reduced in the U.S. due to effective antibiotic therapy [7]. Ulceroglandular tularemia (accounting for more than 90% of cases, especially involving subsp. *holarctica*), often occurs following contact of the skin or mucous membranes by an infected animal or after being bitten by an infected vector, and is characterized by the existence of an ulcerated lesion and regional lymph nodal swelling [36]. The next form is glandular tularemia, which is similar to, and often grouped with, the ulceroglandular form, but lacks an apparent ulcerated site of infection. Oculoglandular tularemia (accounting for 1-4% of cases) results from direct mechanical inoculation of the eye(s), likely by fingers that have handled a contaminated source, and often is characterized by nodules and/or ulcers on the conjunctiva and by regional swollen lymph nodes. Orophyrangeal tularemia (rarely acquired) is the result of ingestion of contaminated food or water, and is often characterized by a severe sore throat with swollen tonsils and cervical lymph nodes, and may occasionally result in death if untreated. Typhoidal tularemia is a term for a severe systemic form of the disease and is apparently associated with subsp. *tularensis*, but the patient lacks the characteristic signs such as lymphadenitis, cutaneous ulcers or lesions, or primary pulmonary involvement. Like the pneumonic form, it may have an untreated mortality rate of 30%-60% [4, 7, 10, 17, 36]. These forms are summarized in table 1-4.

1.1.5 Pathogeneisis and Host-Pathogen Interactions

F. tularensis pathogenicity has been evaluated in most detail using human-avirulent strains in murine models, and therefore such models will be assumed for the purpose of this review. Results from experiments involving human models (i.e., from actual tularemia cases or

human-cell lines) will be specifically referenced for clarification. *F. tularensis* can infect a broad range of cell types, but its primary target appears to be the macrophage [37] as has been demonstrated using primarily the subsp. *holarctica* live vaccine strain (LVS). While murine macrophage models have been widely used to describe *F. tularensis* pathogeneisis, these models do not fully reflect host-cell interactions in humans since the effect of LVS on humans is relatively benign as compared with its effect on mice which are highly sensitive to LVS. In addition, due to concerns of laboratory-acquired infection using subsp. *tularensis*, such correlations have yet to be made using fully virulent subsp. *tularensis* strains. From what has been elucidated from studies using LVS in mice, the innate (T-cell-independent) component of the immune system is primarily involved early (typically within 3 days) following inoculation, which usually occurs through breaks in the skin, but may also occur through ocular, respiratory tract, or gastrointestinal mucous membranes. In studies involving murine models as well human tularemia cases and human cell-line models, T-cell-dependent mechanisms of the adaptive component of the immune system occur later, usually greater than three days following infection (reviewed in REFS. [4, 10, 38, 39]).

Early events post-infection involve ingestion of *F. tularensis* by, and multiplication to high levels within, murine macrophages [37, 40]. The innate immune response appears to vary depending on the type of macrophages infected (reviewed in REF.[10]). In the case of alveolar macrophages, for example, secretion of tumor necrosis factor-alpha (TNF-α) by bacterial-containing macrophages onto natural killer (NK) cells in-turn stimulates the NK cells to produce and feed-back interferon-gamma (IFN-γ) onto infected macrophages, and thus induces bacterial killing. Alternatively, activation of peritoneal macrophages results in nitric oxide (NO) production which facilitates bacterial killing [37, 41, 42]. As previously mentioned for both murine and human models, in the T-cell-dependent mechanism, macrophages present bacterial antigens in MHC-II context to CD4+ lymphocytes which respond by proliferating and secreting TNF-α, IFN-γ, and interleukin-2 (IL-2), thus inducing macrophages to kill their phagocytized

bacteria (reviewed in REF. [4]). Under conditions in which the macrophage phagosome has deteriorated (discussed below), thus resulting in *F. tularensis* residing freely in the cytoplasm, it is likely that antigen presentation occurs via the MHC-I presentation pathway involving CD8+ T-cells [43].

The earliest pathogen-host response appears to be a chemokine (i.e., CXCL8)-mediated recruitment of circulating neutrophils to the surface of *F. tularensis* lipopolysaccharide (LPS)-activated human umbilical endothelial cells (HUVEC) involving E-selectin, VCAM-1, and ECAM-1 interactions [10, 44, 45]. As for the LPS of *F. tularensis*, it may also have a direct role in pathogenesis as will be discussed later. In this initial pro-inflammatory interchange macrophages are not likely involved. One model proposes that this initial influx of neutrophils results in limited *F. tularensis* killing, and that the dead bacteria actually promote the influx of macrophages, which then phagocytize both dead and viable bacterial cells in a cytokine-driven fashion as previously described (reviewed in REF. [10]), and without triggering the respiratory burst [46].

F. tularensis is contained inside a phagosome after entry into the macrophage. Containment within the phagosome has been thought to facilitate iron-dependent bacterial growth due to an acidic pH-associated release of iron from host-cell transferrin [47]. A study by Clemens *et al.*, however, demonstrated that phagosomes of human macrophage-like cells containing live *F. tularensis* in experiments using both LVS and a subsp. *tularensis* isolate only acidified to a pH of 6.7 as compared to 5.5 for phagosomes containing killed *F. tularensis* [43]. According to Oysten *et al.*, this finding casts uncertainty on the actual mechanism for iron acquisition in the absence of acidification [10]. Recently, the complete genome sequence of subsp. *tularensis* strain SCHU S4 revealed genes predicted to encode the ferric uptake regulator (Fur) which, in many other microorganisms, has a key role in modulating iron uptake. This protein and several others encoded by genes identified in SCHU S4 (possibly regulated by Fur itself) may be essential to iron acquisition in *F. tularensis* as well (reviewed in REF. [48]).

Regardless of the exact mechanism of iron uptake, studies of infected mouse and human macrophages using confocal and electron microscopy have shown that *F. tularensis*, by some poorly understood mechanism, is able to escape its phagosome after 3-4 hours of infection [43, 49]. In these studies, after the indicated time most of the bacteria were no longer enclosed by a phagosomal membrane, but were instead free in the cytoplasm where they proceeded to replicate. The cytoplasmic face of the phagosomal membrane apparently acquires a densely staining fibrillar coating which is followed by disintegration of the membrane and liberation of the bacteria into the cytoplasm (reviewed in REF. [10]). It has recently been suggested that this escape is affected by IFN-γ since in treated mouse peritoneal exudate cells the proportion that escaped was significantly lower (80%) than in untreated cells (97%) as determined by transmission electron microscopy (TEM). By contrast, less than 1% of mutant bacteria lacking expression of a 23-kDa protein denoted IglC were able to escape from the phagosome [50]. Within the first 12 hours intracellular-bacterial replication in macrophages is slow, but increases rapidly after that time point, such that host-macrophage apoptosis is initiated [40, 51]. After apoptosis occurs, large numbers of *F. tularensis* are liberated, allowing infection of new cells (reviewed in REF. [10]).

1.1.6 Molecular-Basis of *F. tularensis* Virulence

Currently, the molecular basis of virulence for *F. tularensis* is not well understood, and in particular, the significant subspecies-specific differences in virulence remain a highly pursued area of investigation. This knowledge gap is due in part to limited research with live cultures because of the high risk of laboratory-acquired tularemia as well as a relative lack, until very recently [52, 53], of genetic tools common for other organisms. Despite such limitations, and in part due to bioinformatics resulting from the completed subspecies *tularensis* strain SCHU S4 genome sequencing project [48], some genes potentially associated with virulence have been

identified, and many appear linked to the organism's intracellular growth in macrophages which is necessary to cause disease.

As reviewed by Oysten *et al.* [10], when the proteome of free-living broth-grown *F. tularensis* cells was compared to that of cells grown in a murine macrophage cell line, the production of four proteins was shown to be increased [54]. Although three of the proteins have not been identified, the fourth one, the 23-kDa IglC protein, is also upregulated under oxidative stress conditions, and it has been shown to be part of the intracellular growth locus (*igl*) operon termed *igl*ABCD [55]. As shown by studies in subspecies *holarctica* and *novicida* [56, 57], IglC was shown to be essential in intracellular multiplication in both amobae and murine macrophages [58]. In addition, a study by Telepnev *et al.* suggested that the IglC protein may have a role in inhibiting TNF-α and IL-1 production in infected macrophages, and that in macrophages it may also have a role in Toll-like receptor 4 (TLR4)-mediated signal transduction disruption [57].

The macrophage growth locus (*mgl*)*AB* operon encodes the gene products MglA and MglB, both transcriptional regulators, and from studies involving *F. tularensis* subsp. *novicida,* both appear required for intracellular growth [59]. MglAB, in particular MglA [58], regulates the transcription of several genes including *igl*A, *igl*C, *igl*D, and the pathogenicity determinant protein (*pdp*) genes including *pdpA* and *pdpD* (reviewed in Ref: [10, 48]). In addition, it appears MglAB may also regulate expression of an exported phospholipase C gene, *acp*A, thought to be involved in inhibition of the respiratory burst upon macrophage entry as well as intramacrophage-phagosomal membrane degradation and escape into the cytosol [10, 48, 58, 60] Also recently, a 33.9-kb pathogenicity island (denoted FPI for *Francisella* Pathogenicity Island) has been discovered, and it has been shown to contain the *pdp*A through *pdp*D genes as well as the *igl*ABCD operon [55].

In the work characterizing the FPI, some key findings were made (reviewed in REF: [55]) which improve our understanding of *F. tularensis* virulence. For example, it was shown that transposon-insertion inactivation of the *pdp*A gene diminished intramacrophage growth and

virulence in mice. In addition, it was demonstrated by PCR analysis that the *pdp*D gene amplicon differs in size between subspecies *tularensis* and *novicida*, but that the gene is absent in subspecies *holarctica*, thus making it a strong candidate determinant of subspecies-specific virulence patterns. It was also shown that the *igl*C gene, as well as the entire FPI, is duplicated in LVS [48, 55]. Also noted is that the FPI is surrounded by transposable elements, and therefore it may even be mobile [48, 55]. With regards to transposable elements, the actual discovery of the FPI resulted from bioinformatics-analysis of the genome region containing insertion sequence element (ISE)-mediated mutations in the linked genes, *igl*B and *igl*C, which reduced intramacrophage growth of the mutants [55, 56]. The finding of IS elements in *F. tularensis* is not surprising; and in fact, the recent completed genome sequence of subsp. *tularensis* SCHU S4 revealed that its genome may contain as many as fifty ISFtu1 (IS630 family) elements, sixteen ISFtu2 (IS5 family) elements, three ISFtu3 (ISHpal-IS1016 family) elements, and one each of ISFtu4 (IS982 family) and ISFtu5 (IS4 family) elements [48]. The vast number of IS elements coupled with observations that they have been shown to alter the organism's phenotype as reported by Nano *et al.* [55] and others supports our hypothesis that ISE-mediated insertion/deletion events may contribute significantly to subspecies-level, and even geographic-level divergence and diversification of *F. tularensis*, and may play a role in the diversity of other organism as well.

Some other factors are thought to be involved in *F. tularensis* virulence. The 29-kDa MinD protein has also been reported to be essential for survival in macrophages [61]. The potential requirement of MinD for the maintenance of cell-wall integrity, as well as its role as a heavy-metal ion pump, possibly for radical or toxic ions and which may help the organism resist oxidative killing, are briefly reviewed by Oysten *et al.* [10]. Also discussed in the review [10] was a *val*A gene-encoded ABC transporter possibly required for LPS transport to the *F. tularensis* outer membrane [62]. Studies of a subspecies *novicida* mutant with an inactivated *val*A gene demonstrated that the mutant was unable to grow in macrophages and had an increased

serum sensitivity, which demonstrates a potential role of *F. tularensis* LPS in pathogeneisis [63]. In addition to the potential TLR4-mediated signal transduction disruption properties of IglC as previously mentioned [57], earlier cytokine-induction studies demonstrated that levels of IL-1 and TNF induction in mononuclear cells were dramatically lower for *F. tularensis* LPS as compared with levels for *Escherichia coli* LPS which, like in many other pathogens, acts as an endotoxin capable of inducing these pro-inflammatory cytokines by activating TLRs [10, 64]. This comparatively lower endotoxin property of the *F. tularensis* LPS may therefore be responsible for limiting the initial innate response to neutrophils and avoidance of respiratory burst activation as previously discussed [10, 46], and thereby confer "Trojan Horse" properties on *F. tularensis* to attract unsuspecting macrophages. As a pathogen lacking type III, IV, or V export systems as revealed by the completed SCHU S4 genome sequence, all genes likely encoding a type IV pilus apparatus associated with virulence properties, such as adhesion to host surfaces, in other pathogens [65] has been found instead [48]. The SCHU S4 genome sequencing project also revealed a gene cluster possibly encoding a previously poorly characterized capsule (reviewed in REF: [10]) as well as homologs of the genes, *cap*B and *cap*C, required for biosynthesis of the *Bacillus anthracis* capsule [48] required for full virulence of that organism [66, 67]. A summary of all these known or putative virulence factors or features is presented in table 1-5.

1.2 Molecular Genotyping And Differentiation Methodologies

Whereas the completed SCHU S4 genome sequence has, to date, provided the highest resolution analysis of the organism's composition and organization, the next section will describe, in detail where possible, several of the numerous other molecular methodologies employed thus far to identify and characterize *Francisella* and its representative species and subspecies. As presented in the preceding sections, especially as pertains to molecular determinants of pathogenesis, many of the advancements in our present understanding of *F. tularensis* have

occurred only during the last few years, due largely to the methodologies which follow after a brief section describing them.

1.2.1 Molecular Genotyping Methodology – General Descriptions

1.2.1.1 16S rRNA/rDNA Sequencing:

Sequence analysis of the 16S rRNA gene has provided a highly accurate and versatile method for bacterial classification and identification, even in cases where the organism in question has not been culturable. This methodology has been successfully adapted to polymerase chain reaction (PCR) amplification. As demonstrated by Weisburg *et al.*, the amplification by PCR of a taxonomically diverse collection of eubacterial 16S rDNA genes was possible with a small number of primers; and the resultant amplicons were readily cloned for sequencing or were able to be sequenced directly. The authors' ability to determine rRNA sequences from ATCC lyophilized ampules, without culture, demonstrated that the phylogenetic classification of fastidious or pathogenic species was possible without specialized microbiological methods [68].

1.2.1.2 Pulsed-Field Gel Electrophoresis (PFGE):

PFGE, first described by Schwartz and Cantor in 1984 [69] is a method by which extremely large DNA fragments and plasmids (~30 kb-10,000 kb) can be separated on an agarose gel following restriction enzyme digestion of intact genomic DNA (from agarose plugs containing intact cells which are first lysed to liberate their DNA within the plug) in the same agarose gel by forcing directional changes of the migrating bands during electrophoresis. As a result, different sized DNA fragments are oriented with smaller ones moving in the new direction more quickly than the larger fragments which lag behind. Genotyping using this method can be customized to a given organism based on the restriction enzymes selected in conjunction with the specified separation parameters used.

1.2.1.3 Amplified Fragment Length Polymorphism (AFLP):

This method was introduced by Vos *et al.* [70] and is based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA allowing sets of restriction fragments to be visualized without knowledge of nucleotide sequence, thereby providing a very powerful fingerprinting technique for DNAs of any origin or complexity.

1.2.1.4 Restriction Fragment Length Polymorphism (RFLP):

RFLPs are slight but unique differences observed in the banding patterns of DNA fragments from different individuals of a species when subjected to restriction analysis. Wyman and White have been credited with discovery of the first polymorphic RFLP marker in 1980 [71]. Such differences in RFLP profiles have revolutionized criminal investigations and have become powerful tools in such cases as identifying individuals in paternity cases, population genetics, and diagnosing a variety of diseases. In traditional RFLP genotyping, restriction-digested DNA fragments are generated and then followed by electrophoresis. Southern-blot transfer is performed after electrophoresis, and then the membrane is hybridized using (a) probe(s) containing a sequence of interest, and which if identified, demonstrates the presence of and respective size of the targeted DNA fragment.

1.2.1.5 Polymerase Chain Reaction (PCR):

Recent advances in PCR technology have included the development of TaqMan® and other real-time (RT) fluorescent-based PCR assays. These assays provide increased specificity and test-turnaround (often within one hour) as compared with gel-based assays, and can provide increased sensitivity and low limits of detection (down to femtogram levels), which is especially helpful when the number of organisms is expected to be low. RT-PCR assays facilitate multiplexing of several single-target (singleplex) PCR assays for improved specificity, thereby

decreasing the likelihood of false positive reactions. Also, in part due to the increased resolution (i.e., total genome size and/or numbers of loci analyzed) afforded by multiplexing PCR assays, the designs themselves of which have been derived from improved, higher resolution genetic-target discovery methods such as completed whole genome sequences like *F. tularensis* SCHU S4 [48] and comparative genomic hybridization (CGH) microarrays (which will be discussed later), such PCR strategies are now reasonably differential both at the subspecies and strain level.

Below are examples of PCR assays which have been used for the study of *F. tularensis*, and which will be discussed in more detail later:

1.2.1.5.1 rep-PCR:

Repetitive element sequence-based PCR (rep-PCR) is a group of methods which generate DNA fingerprints that allow discrimination between bacterial strains. Two main sets of repetitive elements are used for typing purposes. The repetitive extragenic palindromic (REP) elements are 38-bp sequences consisting of six degenerate positions and a 5-bp variable loop between each side of a conserved palindromic stem. The enterobacterial repetitive intergenic consensus (ERIC) sequences are another set of DNA sequences which have been successfully used for DNA typing. ERIC sequences are 126-bp elements containing a highly conserved central inverted repeat and are located in extragenic regions of the bacterial genome [72]. On the other hand, the random amplified polymorphic DNA (RAPD) PCR assay is based on the use of short random sequence primers, about 10 to 20 bases in length, which hybridize with sufficient affinity to chromosomal DNA sequences at low annealing temperatures such that they can be used to initiate amplification of regions of the bacterial genome [72]. REP, ERIC, and RAPD sequences have been used as primer binding sites to PCR amplify the genomes of a variety of bacteria.

1.2.1.5.2 VNTR:

Multi-locus variable number tandem repeat (VNTR) analysis (MLVA) is a multiplex-PCR genotyping method based on variable sizes of multiple VNTR loci. VNTRs, also known as short sequence repeats (SSRs) or micro-satellites, are inherently unstable and undergo frequent variation in their repeated units through such mechanisms as slipped-strand mispairing during DNA sysnthesis, and for this reason have been termed "molecular clocks" for monitoring microbial genome evolution [73]. VNTRs have been used for individual strain discrimination within other bacterial species with little genomic variation, i.e., *Bacillus anthracis* and *Yersinia pestis* [74, 75], as well as within smaller collections of North American and Eurasian *F. tularensis* isolates [76, 77].

1.2.1.6 Comparative Genomic Hybridization (CGH) Microarrays:

While the combined successful use of MLVA in the two studies presented [2, 78] demonstrates its utility as a rapid high-resolution subtyping system to understand natural population structures, higher resolution has recently been made possible through use whole-genome CGH microarrays. Besides allowing comparisons at the whole-genome level, CGH microarrays offer the additional advantage of allowing sequence analysis of cloned fragments constituting genomic regions of difference (RD) between the reference and tester strains compared. The last section on molecular methodologies for characterization of *F. tularensis* will focus on two CGH microarray studies.

1.2.2 Molecular Genotyping Methodology – Applications for *F. tularensis* Differentiation

The next section provides detailed findings resulting from molecular genotyping and differentiation of *F. tularensis* using these methodologies. A brief summary of these findings and benefits of each method is provided in table 1-6.

1.2.2.1 Genotyping *F. tularensis* by 16S rRNA/rDNA Sequencing

One of the first molecular methodologies utilized in the characterization of *Francisella* is 16S rDNA sequencing. Application of this methodology for *Francisella* has helped in the phylogenetic classification of the Genus, but with only limited successes of differentiating within the *F. tularensis* species. For example, 16S rDNA analysis has demonstrated the existence of only a few closely-yet-distantly related organisms such as the intracellular pathogen *Wolbachia persica* [79], and as reviewed by Titball *et al.* [15], the fish pathogen *Piscirickettsia salmonis* and other water-associated bacteria such as *Thiomicrospira nivea* and *Cycloclasticus pugitii*, as well as a ciliate endosymbiont, *Caedibacter taeniospiralis*, also appear related to *F. tularensis*. Among pathogens of animals and humans, 16S rRNA sequence analysis has demonstrated that *Coxiella burnetti* and *Legionella* species are most closely related to *Francisella*.

At the *Francisella* genus and species levels, 16S rRNA analysis helped in early molecular identification strategies of the *Francisella* genus from other organisms, and it also provided one of the first molecular methods for differentiating subsps. *holarctica* from *tularensis*, though some cross-reactivity was unavoidable [80]. Work by Sandstrom *et al.* demonstrated that *F. tularensis* strains currently known as subspecies *mediaasiatica* and *holarctica-japonica* share the subspecies *tularensis* 16S rRNA genotype, irrespective of the fact that their virulence and some of their biochemical characteristics conform to those of subspecies *holarctica* genotype strains [25]. Work by Forsman *et al.* also demonstrated limited differential potential of the 16S rDNA analysis methodology by showing that, on the basis of only six nucleotide differences within the rDNA sequenced amplicons, the *F. philomiragia* and *F. tularensis* species could be differentiated; but that overall, all *Francisella* species strains tested still exhibited very high levels (98.5%-99.9%) of similarity even though some of the subspecies appeared distinguishable based on a few of the six nucleotide differences [79]. The same study supported previous findings and helped establish that (formerly *Francisella* species) *novicida* belonged to the *F. tularensis* species, and that they appeared more related to subsp. *tularensis* than subsp. *holarctica* [79, 80]. In addition, 16S

rRNA/rDNA sequence analysis helped in the identification and classification of the previously mentioned Australian isolate [23] as well as the recent detection of potentially novel or diverse *Francisella*-like strains from Houston environmental samples [81]. In a study by Del Blanco *et al.*, 16S rRNA gene sequencing was performed on 42 isolate DNAs from the first Spanish outbreak, and all were identical by that method, all sharing 100% identity with the 16S rRNA sequence of the subsp. *holarctica* strain LVS [33]. In addition, these sequences were compared to those of the second Spanish (waterborne) outbreak [35] and found the latter to have a single nucleotide polymorphism (SNP) with respect to the published LVS sequence [33]. This finding suggests that the two outbreaks may have been caused by two unrelated subsp. *holarctica* strains; but it also seems possible that a sequencing error may have occurred. At any rate, as mentioned previously, further genotyping is not possible due to the unavailability of isolates or DNA from the second outbreak.

In spite of the advantages offered by 16S rRNA/rDNA analysis, it appears that some level of sequencing and/or biochemical testing remains necessary to definitively differentiate *F. tularensis* beyond the species level. As demonstrated in the literature, one strategy commonly employed to increase discrimination is to combine methods, such as was done by Del Blanco *et al.* [33] in which 16S rRNA gene sequencing was combined with two other molecular methods, PFGE and AFLP. In their study, these methodologies were also used to genotype several (n=62) *F. tularensis* strains, including the 42 which were recovered from the first Spanish tularemia outbreak. Additional strains in the study were from France, the Czech Republic, Russia, and the United States. Another method related to PFGE and AFLP due to utilization of restriction digestion of DNA is RFLP, and all three will be briefly discussed in the next three sections.

1.2.2.2 Genotyping *F. tularensis* by PFGE:

In the study, Del Blanco *et al.* initially used eight restriction enzymes, two producing only four bands and four producing more than 30 bands, with neither case being optimal for

differentiation, especially the latter since too many bands is extremely difficult to interpret. *Xho*1 and *Bam*H1 were finally selected and tested against 49 of the strains, including the 42 Spanish outbreak strains. While distinct band patterns following digestion with *Xho*1 were obtained for each of the single *F. philomiragia*, *F. tularensis* subspec. *tularensis*, subspec. *novicida*, and Russian subspec. *holarctica* strains, a single band pattern "B" was obtained for all the Spanish and Czech samples which were additionally digested with *Bam*H1, thus producing 3 additional band patterns. Combining the band patterns from both enzymes resulted in a total of 7 pulsetypes, one each for *F. philomiragia*, *F. tularensis* subspec. *tularensis*, subspec. *novicida*, and the single Russian subspec. *holarctica* sample, and three pulsetypes for the Spanish outbreak strains, namely pulsetypes II, III, and IV. Interestingly, the pulsetype for all three Czech strains was identical to the Spanish pulsetype III [33].

1.2.2.3 Genotyping *F. tularensis* by AFLP:

AFLP is the final method used in the study by Del Blanco *et al.* In the study, all 62 strains were analyzed using four primer pairs: *EcoR*I-T and *Mse*I-T, *EcoR*I-0 and *Mse*I-CA, *EcoR*I-C and *Mse*I-A, and *EcoR*I-A and *Mse*I-C. Besides individual unique profiles generated for *F. philomiragia*, *F. tularensis* subspec. *novicida* strains, two unique profiles, subcluster A1 and A2, were generated for the five subspec. *tularensis* strains tested. From a comparative analysis of the subspec. *holarctica* strains by PFGE and AFLP, all four which produced a PFGE pulsetype II and all three which produced a PFGE pulsetype IV produced an AFLP profile B3. In addition, 35 of the 38 which produced a PFGE pulsetype III produced an AFLP profile B3, whereas the remaining three (all from the Czech Republic) produced an AFLP profile B2 [33].

Although the majority of subspec. *holarctica* samples producing an AFLP profile B3 also produced a PFGE pulsetype III, the comparative results demonstrated ambiguity. For example, the identical PFGE genetic patterns for geographically unrelated strains from Spain and the Czech Republic could be suggestive of a very close epidemiological relationship between them. The

authors concluded that the finding of the Czech and Spanish isolates sharing the same pulsetype but different AFLP profiles (representing ~3% diversity) could be explained by the higher discriminatory power of AFLP over PFGE, which has previously been reported for other organisms [82]. The fact, however, that all PFGE pulsetype II and IV strains produced a single AFLP profile B3 is likewise unsettling, and suggests to this author that neither method alone is definitively differential beyond the subspecies level.

1.2.2.4 Genotyping *F. tularensis* by RFLP:

In the study by Thomas *et al.* [26], RFLP was used to genotype a collection of seventeen epidemiologically unrelated *F. tularensis* isolates. The methodology used in this study involved identification of specific subpopulations of the genomic DNA containing IS elements, specifically ISFtu1 and ISFtu2 due to their high copy numbers in *F. tularensis* as previously discussed. Similar studies based on IS elements have proven to be highly discriminative for typing of other bacterial species including *Mycobacterium tuberculosis* and *Yersinia pestis*, both of which are considered genetically conserved [76, 83, 84]. On the basis of the RFLP patterns for the *F. tularensis* strains tested, all isolates fell into one of five main groups, namely *F. tularensis* subsp. *tularensis*, the attenuated subsp. *tularensis* strain ATCC 6223, subsp. *holarctica*, Japanese subsp. *holarctica*, and subsp. *mediaasiatica*. According to the authors, the findings of this study contrast those involving *M. tuberculosis* and *Y. pestis* since these organisms have been shown to be genetically diverse in terms of both IS element distribution and copy number [83, 85, 86], whereas even despite the diverse geographical origins of the *F. tularensis* strains tested, the distributions of the IS elements were found to be generally stable among isolates of each subspecies, and therefore are not thought to be frequently involved in genome rearrangements. The findings from this study help further support the recommended separate classification of Japanese subsp. *holarctica* isolates as "subsp. *holarctica* biovar japonica" [25] since such isolates consistently grouped separately from other subsp. *holarctica* isolates. Other findings from this

study showed that the copy numbers of ISFtu1 and ISFtu2 in the *F. tularensis* subsp. *mediaasiatica* isolates tested are most similar to the copy numbers, but not the distribution, of these elements in subsp. *tularensis* isolates; and according to the authors, this may suggest that these two subspecies have relatively similar evolutionary histories [26]. Another interesting finding from this study was that the *Pst*I-digested subsp. *holarctica* LVS profile was distinct from that of other subsp. *holarctica* isolates tested, but that the IS element copy number appeared to be similar between both LVS and the other subsp. *holarctica* isolates, which may indicate these elements are conserved within subsp. *holarctica* strains, even though the genome organization may be different.

1.2.2.5 Genotyping *F. tularensis* using PCR

In addition to 16S rDNA PCR and the PCR assays necessary for AFLP, several other PCR strategies have been devised for identification and differentiation of *F. tularensis* from both clinical and environmental sources. Most are gel-based and target single genes as those encoding outer membrane proteins, i.e., *fop*A and *tul*4 [87-89], which are only species-specific. One excellent example of an *F. tularensis* subspecies-differential PCR assay is the RD1 gel-based singleplex assay [90] which was derived following CGH analysis of representative subspecies strains, and which I have used extensively in my own studies (see chapters 3 and 4). The remaining studies using PCR applications to be discussed are multiplex in nature, either in that they were designed that way, or in that multiple singleplex PCR assays were combined to produce a composite PCR profile. The latter case is represented in a study by De La Puente-Redondo *et al.* in which REP-PCR, ERIC-PCR, and RAPD-PCR were compared [32], whereas the former case is represented by MLVA by Johansson *et al.* [2, 76]. Both will be discussed in the following sections.

1.2.2.5.1 Comparison and Combinations of REP-, ERIC-, and RAPD-PCR:

In the study by De La Puente-Redondo *et al.*, DNA from forty-one *F. tularensis* strains (isolated from hares, humans, ticks, and a vole) including 35 from the previously mentioned Spanish outbreak, the three from the Czech Republic, and the one Russian sample previously described [33] and one subsp. *novicida* strain were tested by the three PCR methods. Four distinct profiles were generated for both the REP- and ERIC-PCR methods, whereas RAPD-PCR using an M13 primer produced five distinguishable patterns as compared to seven patterns using the T3-T7 primers. When the four assays were combined together, the 41 strains were divided into 18 distinct "global" groups (designated A–R). Spanish hare isolates belonged to 10 groups (A–J); Czech hare isolates belonged to groups M and P; Spanish human isolates belonged to 6 groups (A, D, G, J, L, O, and N); human strain SCHU belonged to group N; tick isolates belonged to 2 groups (B and Q); the vole strain belonged to group K; and the subsp. *novicida* strain belonged to group R. When the methods were compared individually, RAPD/T3-T7 exhibited the highest discriminating power whereas REP-PCR displayed the lowest degree of discrimination; and different combinations of two or three of the methods produced intermediate degrees of discrimination. According to the authors, this study represents perhaps the first useful epidemiological-typing method of its kind for *F. tularensis*, and chronologically, was a forerunner to the AFLP method [33] previously described. The authors also discussed the significance that, out of 18 global PCR types, four groups (A, D, G, and J) contained both human and hare isolates, which was indicative that hare *F. tularensis* strains are infectious for humans. In addition, all hare strains were divided into twelve groups, indicating the existence of genetic diversity among the strains isolated from hares. Similarly, diversity among the eight human Spanish isolates was demonstrated since they were classified into six groups (all subsp. *holarctica*), and that the human SCHU (the only subsp. *tularensis* isolate studied) strain was clearly distinct (group N) from the others. Finally, according to the authors, this study appears to

have been the first in which subspec. *novicida* (referred to as *F. novicida* in this report) could be clearly differentiated from the other subspecies [32].

1.2.2.5.2 Genotyping *F. tularensis* Isolates using MLVA:

Worldwide Genetic Relationships among *F. tularensis*: Whereas the previous study used a total of four PCR methods to evaluate 41 *F. tularensis* strains primarily from Eurasia, the study by Johansson *et al.* [2] used 25 individual VNTR loci as a high-resolution typing system to establish genetic relationships among 192 globally diverse isolates (including isolates from subsps. *novicida*, *mediaasiatica*, *holarctica-japonica*, *holarctica*, and 45 subsp. *tularensis* strains) across Eurasia and N. America. The results from this study showed a total of 120 individual genotypes among the 192 strains evaluated. The 45 subsp. *tularensis* isolates were grouped into 39 unique genotypes which were further grouped into two distinct clusters, A.I (n=31) and A.II (n=14). The former group was found to include the highly virulent subsp. *tularensis* strain SCHU S4 whereas the latter was found to contain the subsp. *tularensis* type strain, ATCC 6223; and the study also demonstrated Clade A.I to be significantly more diverse than Clade A.II. Overall the subsp. *holarctica* strains (n=132) from multiple geographic locations within N. America and Eurasia were grouped into only 74 genotypes (suggestive that the subsp. *holarctica* strains are less diverse than either A.I or A.II) which were subgrouped into 5 distinct clades, B.I through B.V., the latter of which included all strains from Japan, thus further supporting the argument that Japanese subsp. *holarctica* strains should be a separate subspecies. Interestingly, MLVA demonstrated that only a few genotypes are present among outbreak isolates from either subsp. *tularensis* or *holarctica*, which suggests that only a few distinct *F. tularensis* populations may circulate during an outbreak episode. This agrees with one of the authors' conclusions that *F. tularensis* is predominantly a clonal pathogen, and supports the previous findings [26] suggesting the relative inactivity of IS elements in *F. tularensis* with regard to recombination. Also, MLVA in this study demonstrated a very close genetic relationship between the previously mentioned

subsp. *tularensis* isolates from Slovakia [24] and the laboratory strain SCHU S4. In addition, the authors indicated that North American and Eurasian *F. tularensis* subsp. *holarctica* strains are genetically distinct, with North American strains being slightly more diverse than Eurasian strains; and thus they suggest that this could represent a rare occurrence of spread from the New World to Old World of a human and animal pathogen [2], which supports this author's own hypothesis.

Genetic Comparison of U.S. *F. tularensis* subpopulations: A very recent study of by Farlow *et al.* [78] used the same MLVA methodology to evaluate 161 North American *F. tularensis* isolates, 158 from the United States and 3 from Canada, altogether including 83 subsp. *tularensis*, 72 subsp. *holarctica*, and 6 subsp. *novicida.* As the Johansson *et al.* study, the MLVA typing system provided good genetic resolution, producing a total of 126 unique genotypes. In this case, since only N. American strains were evaluated, only 4 genetic groups were generated: subsp. *tularensis* A.I and A.II., a single subsp. *holarctica* B group, and subsp. *novicida.* MLVA provided near to total discrimination among the A.I. (n=48/G=42) and A.II. (n=35/G=33) strains, whereas the B strains had the poorest genetic resolution (n=72/G=45), and which again suggests the least diversity.

1.2.2.6 Genotyping *F. tularensis* by CGH Microarrays

1.2.2.6.1 The Broekhuijsen *et al.* Study:

In this particular study [90], twenty-seven *F. tularensis* strains from different parts of the world representing the four subspecies, including subsp. *holarctica* strains from Japan, were compared by hybridization to microarray chips constructed using 1,832 clones from a shotgun library of SCHU S4 DNA, which according to the authors represents more than 95% of the *F. tularensis* genome. Chromosomal DNA from the 27 strains tested showed subspecies-specific differential hybridization to certain probes on the microarray. Generally, strains from the same

subspecies formed clusters. Moreover, the cluster analysis indicated that (i) strains belonging to *F. tularensis* subsp. *mediaasiatica* show close genetic similarity to strains of subsp. *tularensis;* (ii) the type strain of *F. tularensis*, ATCC 6223, is distinct from other strains belonging to subsp. *tularensis*; (iii) the Japanese strains cluster separately from the European and American subsp. *holarctica* strains; and (iv) the single representative of subsp. *novicida* showed a unique localization. Many of the differential hybridization probes (DHP) clustered into contiguous genomic RDs. The authors focused further studies on eight RDs found to differentiate between subsps. *holarctica* and *tularensis* strains and which appear to represent deletions of genome content from subsp. *holarctica* with respect to the reference subsp. *tularensis* strain. Interestingly, each RD was either flanked by, or associated with, direct repeat motifs often associated with IS elements. This analysis identified several genes contained within the RDs, but none from the previous "Molecular-basis of *F. tularensis* Virulence" section (see table 5) were identified. One of the RDs, RD-1, was shown to be highly variable among all the *F. tularensis* subspecies, and as previously mentioned, PCR amplification from this region provided the first singleplex PCR assay capable of differentiating all four subspecies including subsp. *holarctica* strains from Japan.

1.2.2.6.2 The Samrakandi *et al.* Study:

This study [91] differed from that of Broekhuijsen *et al.* in that the strain collection used by the former was entirely from North America, but also in that the microarray by the former was constructed from a larger number of clones (n=7,040) from their subsp. *tularensis* shotgun library, representing nearly 4X coverage of the *F. tularensis* genome as compared with only ~1X coverage for the latter. In this particular study, the strain collection consisted of a total of 17 strains from both subsps. *holarctica* and *tularensis*, and a total of 13 subspecies-specific RDs (called $RD_{tularensis}$ because the sequence is present in subsp. *tularensis* but not in subsp. *holarctica*) were observed, five of which were also observed by Broekhuijsen *et al.* Also, three additional

RDs were observed in the subsp. *holarctica* laboratory strain, LVS (called RD_{LVS}). As also seen by Broekhuijsen *et al.*, all RDs were adjacent to transposase-like sequences or repeat sequences which may therefore have facilitated transposition or recombination events during divergence of the subspecies. The authors commented from this study that several genes from among the 13 $RD_{tularensis}$ could contribute to unique virulence or ecological characteristics, including methylases, aminopeptidases, *pdp*-like proteins (previously mentioned in the "Molecular-Basis of *F. tularensis* Virulence" section and shown in table 1-5) and transport proteins. They also commented that of the three RD_{LVS}, though each could encode proteins important in virulence for humans, the most obvious was the Type IV fimbrial protein gene (also previously mentioned in the "Molecular-Basis of *F. tularensis* Virulence" section and shown in table 1-5).

1.3 Hypothesis and Transition to Experiments in Chapters 3 and 4:

The background presented in this chapter raises some interesting questions considering that the genome content between the main *F. tularensis* subspecies, *tularensis* and *holarctica.* Although the vast majority of genomic content is conserved, some of the research characterizing relative pathogenicity and comparative genome studies identified some gene candidates explaining differences in virulence between the two subspecies. Considering the advancements made in genome characterization and differentiation as demonstrated in this chapter, is it possible to identify more gene candidates responsible for the differential virulence of the *holarctica* and *tularensis* subspecies? Given the vast number of wildlife species known to maintain or amplify *F. tularensis*, and the tremendous geographic range from which the organism has been detected, can geographically-associated (phylogeographic) differences within each respective subspecies be identified at the molecular level? As observed from previous studies presented in this chapter, IS-elements are numerous throughout the *F. tularensis* genome and have been implicated in altering gene expression and/or IS-mediated deletion of genetic regions of subsp. *holarctica* relative to subsp. *tularensis.* Based on this observation, my overall hypothesis was that IS elements are the

driving mechanism of divergence between the *tularensis* and *holarctica* subspecies. Based on the enormous ecological domain of *F. tularensis* as well as the preceding hypothesis, I further hypothesized that such IS-mediated phylographic variation exists and can be detected at the molecular level. The experiments presented in chapters 3 and 4 will assess the efficacy of these hypotheses.

Table 1-1. Key *Francisella* Biochemical or Morphological Differential Tests.

As shown in the table, *F. philomiragia* can be differentiated from *F. tularensis* based on a positive oxidase test for the former. The *F. tularensis* subspecies can be further differentiated based on composited results of additional tests including glycerol fermentation, glucose utilization, citrulline ureidase, requirements for supplemental cysteine in growth media, and vegetative cell size. Expected positive results are shown in red.

Table 1-1

Biochemical or Morphological Trait	*Francisella* Genus Differential		*F. tularensis* Species Differential			
	F. philomiragia	*F. tularensis*	Subsp. *tularensis*	Subsp. *holarctica*	Subsp. *novicida*	Subsp. *mediaasiatica*
Oxidase	Pos	Neg				
Glycerol Fermentation	→		Pos	Neg	Pos	Pos
Glucose Utilization	→		Pos	Pos	Pos	Neg
Citrulline Ureidase	→		Pos	Neg	Neg	Pos
Req. Cysteine Supp.	→		Pos	Pos	Neg	Pos
Vegitative Cell Size	→		0.2-0.7 μ	0.2-0.7 μ	0.7-1.7 μ	0.2-0.7 μ

Table 1-2. Terrestrial and Aquatic Tularemia Cycles.

The table shows the source host reservoirs and arthropod vectors associated with both the terrestrial and aquatic cycles of tularemia. As shown, hares and rabbits are commonly the host-reservoir mammals whereas biting flies and ticks serve as the arthropod vectors for the terrestrial tularemia cycle. Beavers, muskrats, and voles serve often as the amplifying host-reservoir mammals whereas mosquitoes and protozoa have been implicated as the arthropod vectors or non-amplifying hosts for the aquatic tularemia cycle.

Table 1-2.

Source	Cycles of Tularemia Infection	
	Terrestrial	Aquatic
Amplifying Host Reservoir	Hares & Rabbits	Beavers, Muskrats, & Voles
Arthropod Vector Non-Amplifying Host	Biting Flies & Ticks	Mosquitoes & Protozoa

Table 1-3. United States Terrestrial Tularemia Cycle Correlation.

The table shows a correlation, primarily of the arthropod vectors, associated with tularemia cases between Western and Central United States. The main difference shown is that, whereas ticks are implicated in *F. tularenisis* transmission for both geographic regions, biting flies have also been implicated in the transmission of *F. tularensis* in Western United States.

Table 1-3.

Geographical Correlation	United States Terrestrial Tularemia Cycle				
	Western States			Central States	
Main Risk Factor	Infected Amplifying Host	Ticks	Biting Flies	Infected Amplifying Host	Ticks

Table 1-4. Clinical Forms of Tularemia.

As shown in the table, there are five main recognized forms of clinical tularemia: Ulceroglandular (frequently grouped with glandular); oculoglandular; orphyrangeal; typhoidal; and pneumonic. The table also shows the respective case frequency, primary cause, and estimated mortality rates for each clinical form. As shown, the typhoidal and pneumonic forms are primarily associated with subsp. *tularensis*; and although their occurrence is rare, they comprise the highest mortality rates, especially when untreated.

Table 1-4.

Form	Clinical Forms of Tularemia		
	Case Frequency	Primary Cause	Mortality Rate
Ulceroglandular/Glandular	>90% (1° association with subsp. *holarctica*)	Contact with infected host or arthropod bite	<3%
Oculoglandular	1-4%	Accidental eye contact from infected source	Not described
Orophyrangeal	Rare	Ingestion of contaminated food or water	Occasional if untreated
Typhoidal	Rare (1° association with subsp. *tularensis*)	Unknown; systemic spread	30%-60% if untreated
Pneumonic	Rare (1° association with subsp. *tularensis*)	Inhallation or septicemic spread from 1° site of infection	30%-60% if untreated

Table 1-5. Known or Putative Virulence Factors.

This table provides a summary of known or putative virulence factors or related features. The known or suspected function or role, as well as the mode of action or observed characteristic of each virulence factor or feature are listed. Macrophages are abbreviated as MΦ.

Table 1-5.

Virulence Factor/Feature	Table of Known or Putative Virulence Factors	
	Function/Role	Mode(s) of Action/Observed Characteristics:
FPI	33.9 Kb region containing several key virulence genes	Entire region duplicated in LVS
		Flanked by transposable elements
		Contains *pdp*A-*pdp*D genes & *igl*ABCD operon
IglC	Required for intacellular multiplication in murine MΦ & amobae	Duplicated in LVS
		Inhibits TNF-α & IL-1
		Disrupts TLR4 signal transduction
PdpA	Required for intra-MΦ growth and virulence in mice	Transposon-insertion inactivation diminishes intra-MΦ growth and virulence in mice
PdpD	Strong candidate for subspecies-specific difference in virulence	The *pdp*D gene is absent in subsp. *holarctica*
		The *pdp*D gene differs in size between subsps. *tularensis* and *novicida*
AcpA	Exported phospholipase C protein	Possibly involved with respiratory burst inhibition upon entry of bacteria into MΦ
		Potentially involved in intra-MΦ-phagosomal membrane degradation and escape into the cytosol
MglA & MglB	Transcriptional regulators	Required for intracellular growth
		Regulates transcription of *igl*A, *igl*C, *igl*D, *pdp*A, *pdp*D, and *acp*A genes
MinD	Essential for survival in MΦ	Possibly required for maintenance of cell-wall integrity
		Possibly serves as a heavy-metal pump for radical or toxic ions to help resist oxidative killing
IS Elements	Numerous throughtout the *F. tularensis* genome	Associated with phenotypic disruption, i.e., PdpA inactivation
		Found associated with flanking all known RDs from CGH studies, possibly implicating them with RD-associated deletion events
ValA	ABC transporter	Apparently required for LPS transport to the *F. tularensis* outer membrane
		Subsp. *novicida* mutant with inactivated *val*A gene was unable to grow in MΦ
LPS	Weakly endotoxigenic in *F. tularensis*	Induces comparatively lower IL-1 & TNF-α in mononuclear cells than *E. coli*
		Helps limit innate immune response and avoid respiratory-burst activation to allow recruitment of unsuspecting MΦ
Type IV Pilus	Assoc. with virulence properties such as host surface adhesion	*F. tularensis* lacks type III, IV, and V export systems of other pathogens
		All genes likely encoding a type IV pilus apparatus present in the complete sequence of SCHU S4
Capsular Gene Cluster & *cap*B/*cap*C	Encode Capsule-associated proteins	Genes possibly encode capsular proteins. Homologs of *cap*B & *cap*C found in complete SCHU S4 sequence: These genes in *Bacillus anthracis* are required for full virulence of that organism.

Table 1-6. Genotyping Procedures for Differentiating *F. tularensis*.

The table lists several of the known *F. tularensis*-differential genotyping methods. For each genotyping procedure, a brief summary of its sample requirements, inclusive methods, degree to which it differentiates, and comments, such as on its benefits or uniqueness, is provided.

Table 1-6.

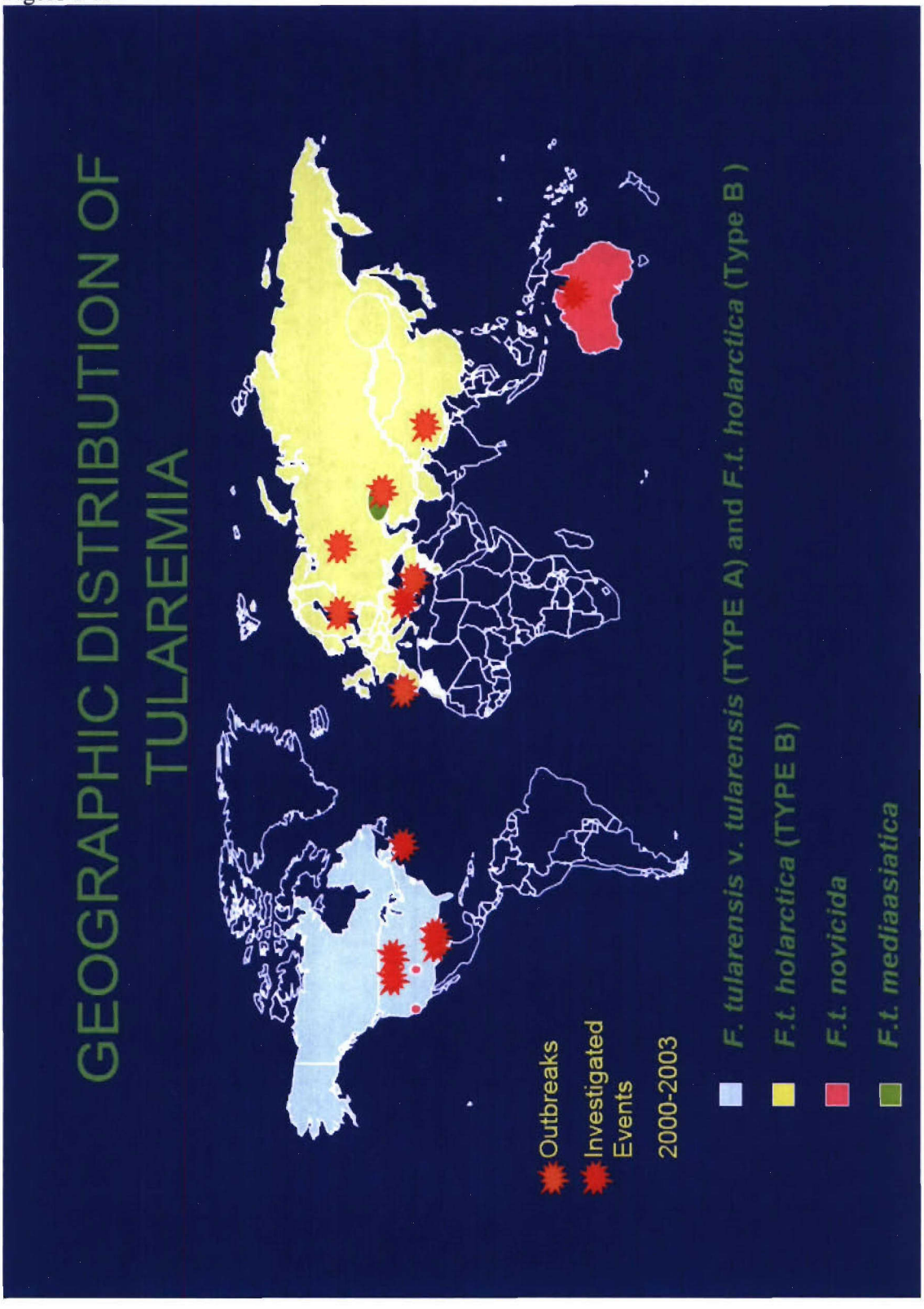
Genotyping Procedure	Required Sample Inclusive Methods	Spec. Diff.	Subsp. Differential					Comments
			tularensis	*holarctica*	*novicida*	*mediaasiatica*	*japonica*	
16S rDNA Analysis	DNA PCR using Universal Primers	Yes	Yes	Yes	Grouped with *tularensis*	Grouped With *tularensis*	Grouped with *tularensis*	Allows Phylogenetic Placement of Unknown Strains Subsp. Differentail Requires Further Testing
PFGE	Whole Organisms Restriction Digestion & Electrophoresis	Yes	Yes	Yes	Yes	Yes	ND	Allows Separation of Large DNA Fragments and Plasmids Limited subsp. Subtyping
AFLP	DNA Restriction Digestion & Kit-Primered PCR	Yes	Yes	Yes	Yes	Yes	ND	Higher Discrimination than PFGE Improved subsp. Subtyping
RFLP	DNA Restriction Digestion, Electrophoresis, Southern Blotting with Custom Probe	Yes	Yes	Yes	Yes	Yes	Yes	Supports subsp. '*japonica*' classification Differentiated LVS from Other subsp. *holarctica* Strains
RD1 PCR	DNA Singleplex PCR with Specific Primers	Yes	Yes	Yes	Yes	Yes	Yes	Insufficient Resolution for subsp. Subtyping
REP, ERIC, & RAPD PCR	DNA Multiplex PCR with Universal Primers	Yes	Yes	Yes	Yes	ND	ND	Good subsp. Subtyping among the *holarctica* Strains
MLVA	DNA Multiplex PCR with Specific Primers	Yes	Yes	Yes	Yes	Yes	Yes	Excellent subsp. Subtyping System Further supports '*japonica*' classification
CGH Microarray	DNA (μg quantities) DNA::DNA Fluoresecent Hybridization	Yes	Yes	Yes	Yes	Yes	Yes	Very High Resolution & Informative Differential Based on Genome Sequence Differences

Figure 1-1.

Map of Geographic Distribution of Tularemia taken from CDC Website [92].

As shown in the map, the data is current as of 2003. The *tularenis*, *holarctica*, and *novicida* subspecies are shown in North America, whereas only subsp. *holarctica* is shown in Europe and Asia, as expected. Also shown is subsp. *mediaasiatica* in Central Asia and subsp. *novicida* in Australia. Not shown is the subsp. *holarctica* variant '*japonica*' in Japan.

Figure 1-1.


GEOGRAPHIC DISTRIBUTION OF TULAREMIA
Outbreaks
Investigated Events
2000-2003
F. tularensis v. tularensis (TYPE A) and F.t. holarctica (Type B)
F.t. holarctica (TYPE B)
F.t. novicida
F.t. mediaasiatica

CHAPTER 2:

Materials and Methods

2.0 Overview:

All experimental materials and methods common to both chapters 3 and 4 are presented at the beginning of this chapter. All experimental materials and methods specifically employed in only chapter 3 (RD_{Spain}) or chapter 4 (Paired End Sequence Mapping, or PESM) are presented in order of experimental progression, first within each chapter, and then from chapter 3 to chapter 4.

2.1 Common Materials and Methods:

2.1.1 Strain and DNA Collection Cultivation and Composition:

All *F. tularensis* cultures presented in this dissertation were propogated on chocolate agar at 37°C in 5% CO_2. Glycerol fermentation of selected isolates (two from France and 3 from Alaska) was determined using Biolog® (Biolog, Inc., Hayward, CA) or as described [91] for those from the Samrakandi *et al.* paper. DNA samples from all isolates were extracted using either a standard large-scale bacterial genomic DNA preparation protocol by Wilson [93], with omission of the CsCl step due to exceptionally high quality DNA without it, or PUREGENE® DNA isolation kits (Gentra Systems, Inc., Minneapolis, MN). All DNAs were subjected to subspecies-differential PCR with primers to $RD_{tularensis}1$ (Broekhuijsen *et al.* [90]), also known as RD1 throughout this dissertation, while some were also tested with c34-5 (Samrakandi *et al.* [91]) primers. Where information is available, a summary of spatial, temporal, host, and other pertinent demographic information, as well as prior subspecies determinations of all strains and/or DNA used in these studies is presented in Table 2-1. Testing of a few strains was omitted with respect to one or the other chapters due to unavailability of the respective DNA at the time of testing.

2.1.2 Conventional Subspecies-specific PCR:

Although the subspecies of most strains in our collection had been previously characterized by various methods including microarray [90, 91], pulsed-field gel electrophoresis (PFGE) [33, 91], and amplified fragment length polymorphism (AFLP) [33], glycerol fermentation was established for only a limited number of the strains [91] plus a few others not previously published; and therefore we established or validated all subspecies using the RD1 PCR assay found to be differential for all four subspecies [90]. Primers for RD1 were the same as published by Broekhuijsen *et al.* [90]. In addition, some of the DNAs were correlated between RD1 PCR and a PCR assay designed from a different RD, here referred to as c34-5, from the Samrakandi *et al.* paper [91]. In designing the c34-5 PCR assay, an additional 1-2 kb of sequence flanking the left and right ends of the RD was added to ensure inclusion of the junctions. Primers for the c34-5 RD PCR assay were designed using Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi), and were as follows for the forward and reverse, respectively: 5'- GAATGGGTATAGTTTTGCCAGAAG-3' and 5'-GTGTTCTAAAAGTATACCTAGCGGATTAAC-3'. The master mix for a single 25 µl reaction for each assay consisted of 5 mM $MgCl_2$ and 160 µM of each dNTP (Idaho Technology, Salt Lake, UT), 500 nM each of forward and reverse primer (Invitrogen, Carlsbad, CA), and 2.5 Units of Platinum Taq (Invitrogen). Each reaction was conducted on 1.5 µl DNA samples prepared as previously described in 2.1.1. Thermocycling conditions were optimized and performed on a Dyad (MJ Research, Reno, NV) thermocycler according to the following cycling parameters: Initial hold at 95°C for 2 min, 30 sec; 30 cycles of 95°C for 30 sec, 64°C for 1 min, and 72°C for 1 min; final extension at 72°C for 5 min; and a final indefinite hold at 4°C. The amplicons were electrophoresed in 0.8% - 1% agarose gels (containing ethidium bromide) run in 1X Tris-acetate-EDTA (TAE) at 85 V for approximately 1.5 hours, and were imaged on a Syngene GeneGenius Imaging Station (Synoptics, Frederick, MD). For quick reference, the primers for both RD1 and

c34-5, as well as RD_{Spain} (discussed below) are tabularized in table 2-2, panel-a; and the conventional PCR master mix recipe and thermocycling conditions for all three assays are listed in tables 2-2, panels-b and -c, respectively.

2.1.3 AFIP Biodefense RT-PCR:

DNA samples from selected strains including those from Alaska, France, and two from Wyoming, as well as from a representative sample from the Spanish outbreak collection were tested with the AFIP's *F. tularensis* species-specific biodefense panel using freeze-dried TaqMan® Real-Time (RT) PCR reagents on the Ruggedized Advanced Pathogen Identification Device (RAPID® – Idaho Technology) according to the manufacturer's instructions.

2.2 RD_{Spain} Polymorphism Materials and Methods:

2.2.1 Microarray for CGH:

The *F. tularensis* microarray is the same as used by Samrakandi *et al.* [91] The array consists of 7040 clones obtained following preparation of a shotgun library of approximately 1 kb sheared fragments obtained by nebulization of subsp. *tularensis* isolate NE-BC410 (reference strain), and which were spotted onto four 40 x 44 feature subarrays using an OmniGrid arrayer (Gene Machines, San Carlos, CA). To compare diversity, 1-2 μg aliquots of the reference strain and test-strain DNA were random primed with CY-5 and CY-3 dye-labeled nucleotides, respectively, with BioPrime DNA labeling kits (Life Technologies, Rockville, MD). Labeling, hybridization, and image analysis were performed as previously described [91, 94]. In essence, addresses hybridized only by reference strain DNA fragments fluoresced as red spots, those hybridized only by test-strain DNA fluoresced as green spots, and those hybridizing by both fluoresced as yellow spots. Hybridized arrays were read with either a ScanArray 5000 (Perkin Elmer, MA) or GenePix 4000B (Axon Instruments, CA) instrument. Fluorescence intensity

ratios of test/reference for each probe address were converted to binary 1 if <2STDEV from the mean and binary 0 if >2 STDEV from the mean. Cluster analysis and sorting of polymorphisms was performed with the MARKFIND program [91, 94] using the Unweighted Pair Group Method with Arithmetic Means (UPMGA) [91, 95]. Addresses showing group-specific patterns of polymorphism were identified using a function of the MARKFIND program which sorts polymorphic characters in the binary strings relative to user-specified groups of taxa [91, 94, 96].

2.2.2 Mapping Regions of Difference:

Clones from addresses corresponding to array probes of interest i.e., demonstrating polymorphisms between reference and test strains, were subjected to DNA sequence analysis using cycle sequencing with labeled T3 and T7 primers. Sequences were aligned into contigs using Sequencher software (Gene Codes, Inc., Ann Arbor, MI) followed by mapping the contigs onto the *F. tularensis* subsp. *tularensis* strain SCHU S4 (also known as SCHU, SchuS4, Schu-4, or Schu 4) genome sequence [48] (or at http://artedi.ebc.uu.se/Projects/Francisella) using Basic Local Alignment Search Tool (BLAST) searches. Resultant RD were confirmed by Southern Blotting (at Dr. Benson's lab) and PCR. Putative proteins were located using NCBI ORF Finder and protein homologies were identified using NCBI BLAST software. Specific gene identities were obtained by analysis of the SCHU S4 genome sequence using the SeqBuilder Module of Lazergene V.6.0. (DNASTAR, Madison, WI).

2.2.3 PCR Designs and Conditions for RD_{Spain}:

2.2.3.1 Conventional RD_{Spain} PCR:

The PCR assay to detect the RD_{Spain} DNA fragment was conducted on 1-1.5 µl DNA samples prepared as previously mentioned in 2.1.1. In designing the RD_{Spain} PCR assay, an additional 1-2 kb of sequence flanking the left and right ends of the RD was added to ensure

inclusion of the junctions. Primers for the RD_{Spain} PCR assay were designed using DNAMAN® software (Lynnon Biosoft), and were as follows for the forward and reverse, respectively: 5'-GTCTTGTTGAGCAAATGCCC-3' and 5'-CGGAGCAGGCTTAAATAGTGA-3'. The master mix for a single 25 µl PCR reaction consisted of 5 mM $MgCl_2$ and 160 µM of each dNTP (Idaho Technology, Salt Lake, UT), 500 nM each of forward and reverse primer (Invitrogen, Carlsbad, CA), and 2.5 Units of Platinum Taq (Invitrogen). Thermocycling conditions were optimized and performed on both a T-Gradient (Biometra, Göttingen, GE) and Dyad (MJ Research, Reno, NV) thermocyclers according to the following cycling parameters: Initial hold at 95°C for 2 min, 30 sec; 30 cycles of 95°C for 30 sec, 64°C for 1 min, and 72°C for 1 min; final extension at 72°C; and a final indefinite hold at 4°C. The amplicons were electrophoresed in 0.8% - 1% agarose gels (containing ethidium bromide) run in 1X TAE at 85 V for approximately 1.5 hours, and were imaged on a GeneGenius Imaging Station (Syngene, Frederick, MD).

2.2.3.2 Cloning and Sequencing the RD_{Spain} DNA Fragment:

PCR amplicon from a Spanish sample, Tu-19, was cloned into electrocompetent *E. coli* host cells of a TOPO TA cloning kit (Invitrogen). Following selection, several colonies were subjected to mini-prep extraction to obtain plasmid DNA, which was then labeled and sequenced using a Licor Sequencing Electrophoresis system. The resultant sequence was BLASTed against both the (then) draft SCHU S4 and draft LVS whole genome sequences (WGSs).

2.2.3.3 Real-Time (RT) RD_{Spain} PCR Design and Conditions:

The design of the assay was based on employment of TaqMan® fluorophore probes which can be detected on the ABI 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster City, CA). In designing the RT-PCR assay, it was important to decrease the amplicon size down to approximately 200 bp as required by the methodology (150 bp is optimal). Primer Express V.2.0 software (Applied Biosystems) was used, in-part (partial manual selection of

primers/probes was required due to the suboptimal nature of the *F. tularensis* genome, particularly the high AT ratio), to design individual assays for both mutant and wild-type strains. The oligonucleotide sequences for the leftward flank are as follows for the forward and reverse primers (Invitrogen), and probe (Applied Biosystems), respectively: 5'-TTTGTTAGGATTTAGTTTTTGTTTACTTATAGGT-3', 5'-ACTGACTCCCTTAGAACCAGAGTCA-3', and Vic-5'-AGTCGATACTAATCAAKAATTTGTTGCACC -3'-Tamra (where K=T or G). The oligonucleotide sequences for the rightward flank are as follows for its forward and reverse primers (Invitrogen), and probe (Applied Biosystems), respectively: 5'-GAAATATTCCATCTCCATCAAAATGC-3', 5'-ATGGTTTAAAGATGACAATAGTAAGTCGA-3', and Fam-5'-ACTACTTTGATTARGCATAAAAGCAAG -3'-Tamra (where R=A or G). Note that Single Nucleotide Polymorphisms (SNPs) were discovered in both probe segments, and therefore the probes were constructed with degenerate bases at the "K" and "R" positions. The Leftward probe contained a T at the "K" position when BLASTed against the LVS sequence, but it contained a G when BLASTed against the SCHU S4 sequence. The Rightward probe contained a G at the "R" position when BLASTed against the SCHU S4 sequence, but it contained an A when BLASTed against the LVS sequence. For clarity, the respective primer and probe sets are listed in table 2.3.3.2a. To maximize throughput, we constructed a duplex assay consisting of the leftward forward primer as the common forward primer, the leftward probe just inside the deletion and flanked by its reverse primer, and the rightward probe outside the right junction and flanked by its reverse primer. This scheme forced expression of only one or the other probe depending on the presence/absence of the RD_{Spain} DNA segment. The master mix for a single 25 µl reaction consisted of a final concentration each of 1X Universal Master Mix (Applied Biosystems), 400 nM of each primer, 600 nM of each probe, and 0.55 U/rxn of Platinum Taq (Invitrogen). Thermocycling conditions were optimized and performed on the ABI 7900HT Fast Real-Time

PCR System (Applied Biosystems) according to the following cycling parameters: Initial hold at 50°C for 2 min; 95°C hold for 10 min; 45 cycles of 95°C for 15 sec, 56°C for 15 sec, and 61°C for 1 min; and a final indefinite hold at 4°C. For quick reference, the primers, probes, master mix recipe, and thermocycling conditions are tabularized in table 2-3, panels a-c, respectively. ABI SDS V.2.1(Applied Biosystems) Software was used to run the instrument and analyze the data.

2.2.3.4 RT-RD_{Spain} PCR Assay Validation:

Ninety (~94% of the AFIP/Univ. Nebraska collection) DNA samples (including all 46 with the RD_{Spain}, 43 W.T., and the AFIP *F. philomiragia*) previously tested by the conventional-RD_{Spain} PCR assay were tested using the RT-PCR assay, and gave results correlating 100% with results from conventional PCR testing. Data analysis for these 90 samples was performed using "Absolute Quantification" mode of the software. Each run required approximately 2.2 hours.

2.2.3.5 RT-RD_{Spain} PCR Testing of a Global Strain Collection:

Finally, a panel of 319 samples from the Keim Genetics Group's global collection of *Francesella* DNA samples was tested using the RT-RD_{Spain} Deletion PCR assay. The composition of the panel is shown in table 2-4 according to geographic sample distribution. Data analysis for these samples was performed using both "Absolute Quantification" and "Allelic Discrimination" modes of the software.

2.2.3.6 Multi-Locus VNTR Analysis (MLVA):

MLVA was performed as described [2] using primers amplifying regions Ft-M1 through Ft-M25. Note: Since MLVA was accomplished at Dr. Fey's laboratory (UNMC), some of the DNA samples tested from the UNMC collection are not included in the combined AFIP-UNMC master collection (table 2-1). All reverse primers were labeled with IRD800 (Li-Cor, Lincoln, NE) and the reactions were resolved on 41 Cm gels using a Li-Cor 4000L automated DNA

sequencer. Control reactions were run using the SCHU S4 and LVS strains [2]. Sizes of MLVA products were predicted from a combination of known SCHU S4 and LVS sizes, the 50-350 bp and 50-700 bp IRD800 size standards (Li-Cor), and a 1 bp sequencing ladder. The data from each strain was tabulated and integrated with MLVA data from the Keim's Genetics Laboratory at Northern Arizona University (NAU) previously run on an ABI 377 System (Applied Biosystems). Data from the two data sets were converted to numeric-bp values and normalized for allele sizes by the NAU Laboratory. Cluster analysis was performed using the NJ/UPGMA algorithm as implemented in PAUP 4.0 beta 10 [97]. Bootstrap analysis was performed using 1,000 iterations of a NJ/UPGMA search.

2.3 PESM Materials and Methods:

2.3.1 Construction of λ phage library (at Dr. Benson's Lab):

Francisella tularensis subsp. *holartica* strain MS304 is a human isolate obtained in 2002 from the State of Missouri. A library was constructed from MS304 genomic DNA by partial digestion with Sau3A1. After optimization of the partial digestion for 10-15 Kb fragments, 7 ug of genomic DNA was digested with Sau3A1 in separate reactions with 0.0625, 0.0312, and 0.0156 units per microgram for 1 hour at 37°C. The partially cut DNA was electrophoresed on a 0.7% agarose gel along with molecular weight markers, and the regions containing 10-15 kb fragments were excised from the gel. The fragments were electroluted, pooled, and then precipitated to concentrate. Size distribution of the gel-purified fragments was confirmed by agarose gel electrophoresis of a small portion of the fragments alongside molecular weight standards. The remaining purified fragments were then ligated into Lambda DASH II Bam HI (Stratagene, La Jolla, CA). The ligations were packaged using Stratagene's Gigapack III Gold Extract according to the manufacturer's recommendations. The packaged phage were titered on Lambda-sensitive XL1-Blue MRA P2, or simply 'P2', *E. coli* host bacteria. The titer from the

packaging was approximately 5 X 10^6 pfu/ml. Library diversity was examined by restriction digestion and DNA sequence analysis of inserts from 10 independent plaques. The library was then amplified once using P2 host bacteria, and DMSO was added to 7% final concentration in the clarified supernatant. The amplified library had a titer of 2.5 x 10^7 pfu/ml. One ml aliquots were stored at –80°C.

2.3.2 Preparation of λ-phage plaques for isolating cloned DNA inserts:

This procedure was essentially performed as described [98], but with minor modifications according to our specific requirements. Unless otherwise indicated, all growth media, buffers, and reagents were prepared according to described protocol [99]. Lambda-sensitive P2 *E coli* cells from the Lambda-DASH II Bam HI kit (Stratagene) were grown and maintained on Luria Broth (LB) agar without antibiotics. For each plaquing experiment, a morning culture of fresh P-2 cells was initiated by inoculating a colony into a sterile Falcon 50 ml conical tube containing 25 ml NZYM Broth (Q-BIOgene, Irvine, CA) with 0.2% maltose (Q-BIOgene), followed by incubation @ 37°C with shaking at 185 rpm for 5-6 hours, or until cells grew to ~OD_{600} of 0.6. Just prior to that time, dilutions of the Lambda phage stock were made in Lambda suspension medium (SM) buffer to a concentration optimized to generate approximately 120-150 plaques per Petri dish for each of 2-3 dishes. Also at that time, the Petri dishes containing bottom agar (Q-BIOgene) were prewarmed to 37°C, and an aliquot of top agarose (Q-BIOgene) was prepared and cooled to ~50°C. Next, in a sterile 15 ml conical tube for each dish, 200 ul of the P-2 cells were gently added and mixed with 100 ul of diluted λ phage, and the mixture was allowed to incubate at 37°C for approximately 20 minutes to allow phage particles to adsorb to cells. Following incubation, the phage-cell suspension was added to the pre-warmed top agarose aliquot and rapidly mixed without production of bubbles, and rapidly poured on top of the bottom agar ensuring rapid and complete coverage without producing 'bumps' in the agar. After a few minutes to allow any

visible condensate to evaporate, the dishes were placed inverted in the incubator for overnight incubation at 37°C.

The next morning, the plates were removed from the incubator and the plaques were evaluated. The plates were wrapped with parafilm and placed inverted in the refrigerator for a minimum of 4 hours and a maximum of 4 days. Plates containing no more than approximately 150 plaques were selected, and only clearly isolated plaques were further processed by direct-PCR amplification (DPA). 96-well PCR plates (Applied Biosystems) pre-loaded with PCR master mix for DPA were loaded by gouging each candidate plaque with a sterile 20 ul pipette tip such that the tip contained visible plaque material, which was then transferred and mixed into a single well of the PCR plate. This process was completed until 94 wells were loaded and processed for PCR (discussed below).

2.3.3 Direct-PCR Amplification (DPA) of Cloned Inserts:

Due to the large size of the cloned DNA fragments, long-range PCR using a *TaKaRa Ex Taq*™ Hot Start master mix kit (TaKaRa Mirus Bio, Madison, WI) was designed and optimized using T3 (5´-AATTAACCCTCACTAAAGGG-3´) and T7 (5'-TAATACGACTCACTATAGGG-3') primers for amplification initiation at the corresponding Lambda insert-flanking region promoter sites. For each PCR run of 94 reactions, a 98 X master mix was prepared according to the manufacturer's recipe with each single 25 ul reaction containing 500 nM of each primer. All 94 wells of the PCR plate were loaded with master mix using a multichannel pipette (Note: Multichannel pipetting was performed for all subsequent manipulations and transfers involving multiple samples/multi-well reaction plates throughout all experiments). Thermocycling conditions were optimized and performed on a Dyad (MJ Research) thermocycler according to the following cycling parameters: Initial hold at 95°C for 2 mins, 30 sec; 36 cycles of 95°C for 50 sec, 55°C for 50 sec, and 72°C for 15 mins; final extension at 72°C for 5 mins; and a final indefinite hold at 4°C.

2.3.4 PCR Cleanup and Clone-Amplicon Storage:

Following each PCR run, clone-amplicon purification was accomplished using Montage® $PCR_{\mu 96}$ Plates (Millipore, Billerica, MA) which were processed on a SAVM 384 Vacuum Manifold (Millipore) according to the manufacturer's instructions. The final elution was performed using 30 ul of Invitrogen Distilled DNase-/RNase-free water (which was used exclusively throughout all experiments) to ensure sufficient DNA volumes for sequencing and sizing experiments, and all plates were sealed with adhesive sealing lids (Bio-Rad, Hercules, CA) and stored at 4°C. Prior to further processing, each PCR reaction plate was centrifuged briefly in an Eppendorf multi-well plate-spinning centrifuge (Eppendorf, Westbury, NY) to sediment down potentially contaminating DNA droplets, and the sealing lids were then carefully removed to prevent well-to-well carryover. Upon completion of each procedure, new sealing films were applied to each plate, and the plates were returned to 4°C.

2.3.5 Clone-Amplicon Sizing Experiments:

Amplicon-size determinations were performed using agarose gel electrophoresis. 15 cm x 25 cm 0.65% agarose gels containing 51-wells were made in 1x Tris-acetate-EDTA (TAE). Each gel accommodated 47 clone-amplicon samples (5 ul sample plus 3 ul loading dye was added into each lane) and 4 lanes of Bio-Rad 1-15 kb Molecular Ruler (8 ul was added for each lane) for band-size standardization. Once all lanes were loaded, each gel was electrophoresed at 85 V for approximately 4.5 hours. When electrophoresis was completed, each ethidium bromide-stained gel was photographed using a Syngene GeneGenius (Synoptics) imaging system, and the image was analyzed using the GeneTools (Synoptics) software package. The band sizes for all clones and DNA standards were exported from GeneTools into a Microsoft Excel spreadsheet for each run for further analysis.

2.3.6 Clone-Amplicon Paired-End Sequence Determinations:

DNA sequencing reactions for all clones were carried out using BigDye® Terminator v3.1 Cycle Sequencing Kits (Applied Biosystems) with pGEM DNA serving as the sequence reaction controls. Each sequence reaction experiment was setup in 96-well reaction plates, with each sample being divided into two reactions – one reaction with T3 primer and the other with T7 primer (both primers the same as from the initial PCR reactions). pGEM was likewise run as two independent reactions – one reaction with M13-F primer and the other with T7. Sequence reactions were carried out on the Dyad (MJ Research) thermocycler, and each sequence reaction plate was cleaned-up using a Montage SEQ_{96} Kit on the SAVM 384 Vacuum Manifold (Millipore). The labeled DNA was then transferred into a new reaction plate and loaded onto an ABI 3100 automated capillary-electrophoresis (CE) sequencer (Applied Biosystems) which was configured with an 80 cm capillary array and loaded with Performance Optimized Polymer (POP)-4 (Applied Biosystems). Once each sequence run was completed, the pGEM control sequences were evaluated to ensure the sequence reaction and sequence run were successful, and if so, all sample sequence ".abi" files were opened and trimmed in Sequencher V.4.0.5 (Gene Codes), merged, and output as a single FASTA file for further analysis.

2.3.7 Fragment Length and Paired-End Size Sequence Pipeline:

The steps, up to this point, of the PESM protocol are highlighted in figure 2-1. The sizing and sequence files were input into a Perl-based program, referred to as the Paired-End Sequence Mapping program, or rather "pipeline" (PESMP), to identify homologous regions within the draft *F. tularensis* live vaccine strain (LVS) whole genome sequence (WGS) found at (ftp://bbrp.llnl.gov/pub/cbnp/F-tularensis/F.tularensis.html) as well as the completed SCHU S4 WGS [48]. The PESMP program as well as each WGS are contained on the University of Nebraska-Lincoln (UNL) local Pathogene (http://pathogene.unl.edu) web server. The pipeline input data consisted of one trimmed FASTA sequence file and its corresponding tab-delimited

Excel sizing file. By using the tab-delimited sizing file, the program was able to output the corresponding gel-size of each clone-amplicon along with the size determined from the coordinates of each queried WGS. The output for both LVS and SCHU S4 were copied and pasted from the Perl-program into an Excel spreadsheet and then sorted according to the paired-end coordinates of LVS since it is the subspecies *holarctica* strain-type, and our hypothesis was that it should share a high degree of identity with our subspec. *holarctica* library strain, MS-304. By having both the WGS coordinate-based sizes as well as the actual gel size for each clone, we further sorted the data based on agreement, within ~ +/- 2 kb, between the actual and WGS-predicted sizes. Clones having agreement between the gel and both WGS sizes were considered to be *F. tularensis* species-specific, or 'species-conserved', whereas those clones that agreed only with LVS but not SCHU S4 were considered subspecies-specific, such that the size differences corresponded to sequence differences which could potentially account for some observed phenotypic differences between the two subspecies. Following these initial sorting schemes, all clones classified as subspecies-specific were further sorted and those containing sequence within overlapping LVS coordinates were contig'd together. Since these new LVS contigs are representative of the *holarctica* subspecies, they are now called *holarctica*-contiguous regions (CR) or $CR_{holarctica}$. All $CR_{holarctica}$ sequences were saved as text files and used for fine-structure genome mapping.

2.3.8 Fine-Structure Genome Mapping:

Each $CR_{holarctica}$ sequence was BLASTed against the SCHU S4 WGS found on the UNL Pathogene BLAST server. The corresponding BLAST output was saved as a text file, and the resultant subject and query coordinates for all segments/subsegments, except for multiple repeats of IS elements (i.e., ISftu1 & ISftu2 which repeat up to 50 and 16 times, respectively, throughout the *F. tularensis* genome [48]), were entered into an Excel spreadsheet. For each $CR_{holarctica}$, the corresponding SCHU S4 segments/subsegments were then mapped from smallest to largest

sequence coordinates according to the SCHU S4 WGS. Line images for each $CR_{holarctica}$ and its corresponding SCHU S4 map (not drawn to scale) were generated in Microsoft Powerpoint using different colors for different subsegment regions, with same colors showing shared identity between both genomes, and arrows to illustrate sequence synteny and/or order changes in the corresponding SCHU S4 map. In addition, the $CR_{holarctica}$ maps were assembled using the SeqBuilder module of Lazergene V.6.0 (DNASTAR, Madison, WI) to mine out the SCHU S4 genome content from the completed SCHU S4 sequence, and each gene/pseudogene from the corresponding annotation [48] was entered into an Excel spreadsheet. This data was next overlaid on top of the LVS segments (but not the SCHU S4 maps due to scale-limited space constraints) to demonstrate comparative genome structure and order between the two genomes. Putative virulence genes and/or biochemical-associated genes are spelled-out in red on their respective maps. Also, the full names of genes found to be truncated either due to the CR sequence beginning/ending in a gene or due to an INDEL or translocation/inversion event were listed above their respective $CR_{holarctica}$, or rather LVS, segment with a "T" designating the truncation.

For "truer-scale" comparisons, TIGR in-house Perl scripts were run on a Linux platform to generate graphical representations of several of the $CR_{holarctica}$ shown in Fig. 4-6 through Fig. 4-7. The final figures as shown were assembled in Adobe Illustrator 10.

The SCHU S4 sequence of all genes apparently truncated due to a start or termination of a $CR_{holarctica}$ segment, or rearrangement within a $CR_{holarctica}$ segment, were BLASTed against the Pathogene server's LVS WGS for determination of homology between the two respective gene sequences.

2.3.9 Comparative Genome PCR (CG-PCR):

To test our hypothesis that subspecies-specific CRs are in fact different between subsps. *holarctica* and *tularensis* but yet conserved among other strains within each respective

subspecies, a tri-primer nested-PCR assay was designed for each CR based on the bioinformatics analysis used to construct the *F. tularensis* subspecies comparative genome maps. Primers for all assays were designed using Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). The assays were designed by first designating a primer common to both LVS and SCHU S4, either forward or reverse (designated C-F or C-R), immediately adjacent to a breakpoint in SCHU S4 where either synteny of the next segment changed or was translocated leaving a SCHU S4-specific region for a SCHU S4-specific primer, and which likewise left a target for an LVS-specific primer in the adjacent-contiguous LVS sequence. Since either intact or truncated IS elements or their corresponding repeated elements were present at all breakpoints, care was taken to avoid placement of primers into sequence encoding them, and BLAST searches were performed for all primers to limit their placement to their intended location (see table 2-5 for all primer coordinates, sequences, and intended amplicon sizes). All PCR assays were conventional by design and conducted on 1.5 µl of the DNA samples used for RD1 PCR.

All CG-PCR reactions were performed in 25 μl volumes, each containing 5 mM $MgCl_2$ and 160 μM of each dNTP (Idaho Technology), 500 nM each of common primer, LVS-specific primer, and SCHU S4-specific primer (Invitrogen), and 2.5 Units of Platinum Taq (Invitrogen) Thermocycling conditions were optimized and performed on a Dyad (MJ Research) thermocycler according to the following cycling parameters: Initial hold at 95°C for 2 min, 30 sec; 32 cycles of 95°C for 30 sec, 60°C for 1 min, and 72°C for 1 min; final extension at 72°C for 5 min; and a final indefinite hold at 4°C. Each assay was first tested against SCHU S4 and LVS, and then against a 91-strain global *Francisella* DNA panel composed of DNAs from 1 *F. philomiragia*, 3 subsp. *novicida*, 1 subsp. *holarctica-japonica*, and from a combination of 85 spatially and temporally diverse strains representing both the *holarctica* and *tularensis* subspecies (see figure 4-5 for actual CG-PCR panel composition). The amplicons were electrophoresed on 0.8% - 1% agarose gels run in 1X TAE at 85 V for approximately 2 hours, or until adequate size-discrimination was

accomplished. A 100-bp PCR Molecular Ruler ranging from 100 bp to 3 kb (Bio-Rad) was used for size determinations. All ethidium bromide-stained gels were imaged on a GeneGenius Imaging Station (Syngene). All PCR reactions producing negative results (no band seen by gel electrophoresis) or results inconsistent with a strain's known subspecies (as determined by the original contributor and verified by RD1 PCR as previously discussed) were repeated for verification of the initial result.

Table 2-1.

Master Table of *Francisella* Strains Used in Studies.

The table includes the conversion number (or coded name) of each strain or DNA, its collection location, subspecies or species (if not *F. tularensis*), geographic origin, year of isolation, host or vector, and subspecies RD1 PCR result. The original source location for each strain is provided by code, and the key is provided on the last page of the table.

Table 2-1: Master *Francisella* Strain-DNA Collection Part 1.

Ft Panel #:	Strain ID # used in Study	Collection Location	Source*	Species or subspecies	Geographic origin, year of isolation	Host or Vector	Subsp. RD1 PCR
1	Tu-30	AFIP	AFIP 1	*tularensis*	Utah, 1920	Human	Type-A
2	SchuS4	AFIP	CAPM 1	*tularensis*	Strain SchuS4, Ohio, 1941	Human	Type-A
3	A88R160	AFIP	AFIP 2	*tularensis*	Strain A88R160, USA	Rabbit	Type-A
4	LVS	AFIP	USAMRIID	*holarctica*	Russia, 1961	Unknown	Type-B
5	Tu-1	AFIP	ALG1	*holarctica*	Valladolid, Spain, 1997	Hare	Type-B
6	Tu-2	AFIP	ALG2	*holarctica*	Valladolid, Spain, 1998	Hare	Type-B
7	Tu-3	AFIP	ALG3	*holarctica*	Valladolid, Spain, 1998	Hare	Type-B
8	Tu-4	AFIP	ALG4	*holarctica*	León, Spain, 1998	Hare	Type-B
9	Tu-5	AFIP	ALG5	*holarctica*	Palencia, Spain, 1998	Hare	Type-B
10	Tu-6	AFIP	ALG6	*holarctica*	León, Spain, 1998	Hare	Type-B
11	Tu-7	AFIP	ALG7	*holarctica*	León, Spain, 1998	Hare	Type-B
12	Tu-8	AFIP	ALG8	*holarctica*	Palencia, Spain, 1998	Hare	Type-B
13	Tu-9	AFIP	ALG9	*holarctica*	Palencia, Spain, 1998	Hare	Type-B
14	Tu-10	AFIP	ALG10	*holarctica*	Valladolid, Spain, 1998	Hare	Type-B
15	Tu-11	AFIP	ALG11	*holarctica*	Valladolid, Spain, 1998	Hare	Type-B
16	Tu-12	AFIP	ALG12	*holarctica*	Zamora, Spain, 1998	Hare	Type-B
17	Tu-13	AFIP	ALG13	*holarctica*	Zamora, Spain, 1998	Hare	Type-B
18	Tu-14	AFIP	ALG14	*holarctica*	Palencia, Spain, 1998	Hare	Type-B
19	Tu-15	AFIP	ALG15	*holarctica*	Valladolid, Spain, 1998	Hare	Type-B
20	Tu-16	AFIP	ALG16	*holarctica*	Valladolid, Spain, 1998	Hare	Type-B
21	Tu-17	AFIP	ALG17	*holarctica*	Segovia, Spain, 1998	Hare	Type-B
22	Tu-18	AFIP	ALG18	*holarctica*	Palencia, Spain, 1998	Hare	Type-B
23	Tu-19	AFIP	ALG19	*holarctica*	Palencia, Spain, 1998	Hare	Type-B
24	Tu-20	AFIP	ALG20	*holarctica*	Palencia, Spain, 1998	Hare	Type-B
25	Tu-21	AFIP	ALG21	*holarctica*	Palencia, Spain, 1998	Hare	Type-B
26	Tu-22	AFIP	ALG22	*holarctica*	Valladolid, Spain, 1998	Hare	Type-B
27	Tu-28	AFIP	CAPM2	*holarctica*	Strain 130, Czech Republic	Hare	Type-B
28	Tu-29	AFIP	CAPM3	*holarctica*	Strain 2713, Czech Republic	Hare	Type-B
29	Tu-35	AFIP	CAPM4	*holarctica*	Strain T-1/59, Czech	Hare	Type-B
30	Tu-36	AFIP	LEO1	*holarctica*	Palencia, Spain, 1998	Hare	Type-B
31	Tu-37	AFIP	LEO2	*holarctica*	Valladolid, Spain, 1998	Hare	Type-B
32	Tu-38	AFIP	LEO3	*holarctica*	Soria, Spain, 1998	Hare	Type-B
33	Tu-39	AFIP	LEO4	*holarctica*	Zamora, Spain, 1999	Hare	Type-B
34	Tu-40	AFIP	LEO5	*holarctica*	Zamora, Spain, 1998	Hare	Type-B
35	Tu-44	AFIP	LEO6	*holarctica*	Ávila, Spain, 1998	Hare	Type-B
36	Tu-45	AFIP	LEO7	*holarctica*	Valladolid, Spain, 1998	Hare	Type-B
37	Tu-47	AFIP	LEO8	*holarctica*	León, Spain, 1998	Hare	Type-B
38	Tu-48	AFIP	LEO9	*holarctica*	Palencia, Spain, 1998	Hare	Type-B
39	Tu-31	AFIP	ALG23	*holarctica*	Valladolid, Spain, 1998	Human	Type-B
40	Tu-32	AFIP	ALG24	*holarctica*	Valladolid, Spain, 1998	Human	Type-B
41	Tu-33	AFIP	ALG25	*holarctica*	Valladolid, Spain, 1998	Human	Type-B
42	Tu-34	AFIP	ALG26	*holarctica*	Palencia, Spain, 1998	Human	Type-B
43	Tu-24	AFIP	HLE1	*holarctica*	León, Spain, 1998	Human	Type-B
44	Tu-25	AFIP	HZA1	*holarctica*	Zamora, Spain, 1998	Human	Type-B
45	Tu-26	AFIP	HZA2	*holarctica*	Zamora, Spain, 1998	Human	Type-B
46	Tu-27	AFIP	HZA3	*holarctica*	Zamora, Spain, 1998	Human	Type-B
47	AFIP 3	AFIP	AFIP 3	*holarctica*	Chateneaux, France	Human	Type-B
48	AFIP 4	AFIP	AFIP 4	*holarctica*	St. Germaine, France	Human	Type-B
49	Tu-23	AFIP	ALG27	*holarctica*	Zamora, Spain, 1998	Vole	Type-B
50	Tu-41	AFIP	LEO10	*holarctica*	Zamora, Spain, 1998	Tick	Type-B
51	Tu-46	AFIP	LEO11	*holarctica*	Valladolid, Spain, 1998	Tick	Type-B
52	Tu-42	AFIP	CAPM5	*holarctica*	Strain 503, Russia	Tick	Type-B
53	*F.t.n* 15482	AFIP	CAPM6	*novicida*	ATCC 15482, Utah, 1950	Water	*novicida*
54	Tu-43	AFIP	AFIP 7	*novicida*	Texas, 1991	Human	*novicida*
55	D2005067002	AFIP	USAMRIID	*novicida*	Unknown	Unknown	*novicida*
56	*F.ph* 25015	AFIP	FOA1	*F. philomiragia*	ATCC 25015, Utah, 1959	Muskrat	Neg
57	99A-2628	AFIP	CDHS1	*holarctica*	Strain 99A-2628, California	Human	Type-B
58	89A-7092	AFIP	CDHS3	*holarctica*	Strain 89A-7092, California	Squir. Monkey	Type-B

Table 2-1: Master *Francisella* Strain-DNA Collection Part 2.

Ft Panel#:	Strain ID # used in Study	Collection Location	Sourcea	Species or subspecies	Geographic origin, year of isolation	Host or Vector	Subsp. RD1 PCR
59	Japanese	AFIP	AFIP-Jap	*holarctica*- japan	Japan	Unknown	*japonica*
60	Austrian	AFIP	AFIP-Aus	*holarctica*	Vienna	Unknown	Type-B
61	FSG-1	AFIP	NCDC	*holarctica*	Georgia, FSU, 1987	Tick	Type-B
62	FSG-2	AFIP	NCDC	*holarctica*	Georgia, FSU, 1980	Tick	Type-B
63	FSG-3	AFIP	NCDC	*holarctica*	Georgia, FSU, 1977	Bird	Type-B
64	FSG-4	AFIP	NCDC	*holarctica*	Georgia, FSU, 1997	Tick	Type-B
65	FSG-5	AFIP	NCDC	*holarctica*	Georgia, FSU, 1974	Comn. Shrew	Type-B
66	FSG-6	AFIP	NCDC	*holarctica*	Georgia, FSU, 2002	Vole	Type-B
67	FSG-7	AFIP	NCDC	*holarctica*	Georgia, FSU, 2002	Tick	Type-B
68	FSG-8	AFIP	NCDC	*holarctica*	Georgia, FSU, 1997	Vole	Type-B
69	FSG-9	AFIP	NCDC	*holarctica*	Georgia, FSU, 1990	Vole	Type-B
70	FSG-10	AFIP	NCDC	*holarctica*	Georgia, FSU, 1956	Gerbil	Type-B
71	88R52	AFIP	AFIP	*tularensis*	Strain 88R52, USA, 1988	Rabbit	Type-A
72	88R144	AFIP	AFIP	*tularensis*	Strain 88R144, USA, 1988	Rabbit	Type-A
73	AK-1133496	AFIP	AKPHL	*tularensis*	Fairbanks, Alaska, 2003	Arctic Hare #1	Type-A
74	AK-1100558	AFIP	AKPHL	*tularensis*	North Pole, Alaska, 2004	Arctic Hare #2	Type-A
75	AK-1100559	AFIP	AKPHL	*tularensis*	North Pole, Alaska, 2004	Arctic Hare #2	Type-A
76	FR-LR	AFIP	CHUNF	*holarctica*	Lorraine, France, 1993	Human	Type-B
77	FR-SS	AFIP	CHUNF	*holarctica*	Near Langres, France, 2000	Human	Type-B
78	UNMC061598	UNMC	NMC	*tularensis*	NE Ref Strain, Nebraska	Human	Type-A
79	UNL091902	UNMC	UNVDL	*tularensis*	Nebraska, USA	Human	Type-A
80	WY-WSVL01	UNMC	WSVL	*holarctica*	Wyoming, USA	Bovine	Type-B
81	WY-9868529	UNMC	WSVL	*holarctica*	Wyoming, USA	Guinea Pig	Type-B
82	WY-00W4114	UNMC	WSVL	*tularensis*	Wyoming, USA	Prairie Dog	Type-A
83	WY-96194280	UNMC	WSVL	*holarctica*	Wyoming, USA	Rabbit	Type-B
84	WY-WSVL02	UNMC	WSVL	*tularensis*	Wyoming, USA	Human	Type-A
85	OK-00101504	UNMC	OSU	*tularensis*	Oklahoma, USA	Feline	Type-A
86	OK-98041035	UNMC	OSU	*tularensis*	Oklahoma, USA	Feline	Type-A
87	MS-304	UNMC	MPHL	*holarctica*	Missouri, USA	Human	Type-B
88	NC-54558-01	UNMC	RADL	*tularensis*	North Carolina, USA	Feline	Type-A
89	NC-52797-99	UNMC	RADL	*tularensis*	North Carolina, USA	Rabbit	Type-A
90	NC-54559-01	UNMC	RADL	*tularensis*	North Carolina, USA	Feline	Type-A
91	CDC NE 031457	UNMC	CDC	*tularensis*	Lincoln, NE, USA, 2003	Human	Type-A
92	UNL072704	UNMC	UNVDL	*tularensis*	Lincoln, NE, USA, 2004	Rabbit	Type-A
93	ATCC-6223	AFIP	AFIP-6223	*tularensis*	Utah, 1920	Human	Type-A
94	No Code	UNMC	AFIOH	*tularensis*	North Carolina, USA	Rabbit	Type-A
95	MO MS1349	UNMC	MPHL	*tularensis*	Missouri, USA	Human	Type-A
96	AFIOH Feline	UNMC	AFIOH	*tularensis*	Oklahoma, USA	Feline	Type-A
97	MO No Code	UNMC	MPHL	*holarctica*	Missouri, USA	Human	Type-B

a.) Source Legend

AFIOH = Air Force Institute for Operational Health

AFIP = Armed Forces Institute of Pathology, Washington, DC

AKPHL = Alaska Public Health Laboratory

ALG = Central Laboratory of Animal Health, Algete, Madrid, Spain

CAPM = Collection of Animal Pathogenic Microorganisms, Brno, Czech Republic

CDC = Centers for Disease Control, Ft. Collins, CO

CDHS = California Department of Health Services, Sacramento

CHUNF = Lab de Bactériologie, Centre Hospitalier et Universitaire, Nancy, France

FOA = National Defence Research Establishment, Umeå, Sweden

HLE = Dept. of Med. Microbiology, Hospital Princesa Sofia, Insalud, León, Spain

HZA = Laboratory of Microbiology, Hospital Virgen de la Concha, Insalud, Zamora, Spain

LEO = Central Laboratory of Animal Health, León, Spain

MPHL = Missouri State Public Health Laboratory

NCDC = National Center for Disease Control, Tbilisi, Georgia (FSU)

NMC = Nebraska Medical Center, Omaha

OSU = Oklahoma State University

RADL = Rollins Animal Diagnostic Laboratory, North Carolina

UNMC = University of Nebraska Medical Center

UNVDL = Univ. of Nebr. Veterinary Diagnostic Laboratory, Lincoln

USAMRIID = U.S. Army Research Institute of Infectious Diseases

WSVL = Wyoming State Veterinary Laboratory

Table 2-2. Conventional PCR Primers, Master-Mix Recipe, and Reaction Conditions.

Panel 2-2a shows the primers for the RD_{Spain}, and both the c34-5 and RD1 subspecies-differential PCR reactions. Panel 2-2b shows the master mix recipe for each assay. Panel 2-2c shows the thermocycling conditions for each assay.

Panel 2-2a: Conventional PCR Primers.

RD_{Spain} PCR Primers:
Fwd Primer: 5'-GTCTTGTTGAGCAAATGCCC-3'
Rev Primer: 5'-CGGAGCAGGCTTAAATAGTGA-3'
c34-5 Primers:
Fwd Primer: 5'- GAATGGGTATAGTTTTGCCAGAAG-3'
Rev Primer: 5'- GTGTTCTAAAAGTATACCTAGCGGATTAAC-3'
RD1 Primers:
Fwd Primer: 5'- TTTATATAGGTAAATGTTTTACCTGTACCA -3'
Rev Primer: 5'- GCCGAGTTTGATGCTGAAAA -3'

Panel 2-2b: Conventional PCR Master-Mix Recipe.

Components:	Source:	Vol (ul)	Final [ ]
DI H_2O	Invitrogen (Carlsbad, CA)	18	1X
10X 50mM Mg 10X Buffer	Idaho Technology (Salt Lake, UT)	2.5	1X
10X dNTPs	Idaho Technology (Salt Lake, UT)	2	160uM each dNTP
Fwd Primer (25 μM)	Invitrogen (Carlsbad, CA)	0.5	500 nM
Rev Primer (25 μM)	Invitrogen (Carlsbad, CA)	0.5	500 nM
Platinum Taq (5 U/μl)	Invitrogen (Carlsbad, CA)	0.5	2.5 U/RXN
Sample Template DNA	DNA Collection	1	(1-50 ng/RXN)
Total Volume		25	

Panel 2-2c: Conventional Assay Cycling Conditions.

Hot Start	PCR			Final	Final Hold
	Cycle (30 cycles)				
Hold	Denature	Anneal	Extend	Extend	
95°C	95°C	64°C	72°C	72°C	4°C
2 min, 30 sec	30 sec	1 min	1 min	5 min	Infinity

Table 2-3. RT-RD_{Spain} PCR Primers, Master-Mix Recipe, and Thermocycling Conditions.

Panel 2-3a shows the primers,, panel 2-3b shows the master-mix recipe, and panel 2-3c shows the thermocycling conditions. Tm refers to the melting temperature of the each primer or probe.

Panel 2-3a: RT-RD_{Spain} PCR Primers & Probes.

Flank/Junction-1(Left):
Fwd Primer: 5'-TTTGTTAGGATTTAGTTTTTGTTTACTTATAGGT-3' (outside deletion) (Tm=60)
Rev Primer: 5'-ACTGACTCCCTTAGAACCAGAGTCA-3' (inside deletion) (Tm=59)
Probe: Vic-5'-AGTCGATACTAATCAAKAATTTGTTGCACC -3'-Tamra (inside deletion) (Tm=72) *K=T or G
Flank/Junction-2 (Right):
Fwd Primer: 5'-GAAATATTCCATCTCCATCAAAATGC-3' (inside deletion) (Tm=63)
Rev Primer: 5'-ATGGTTTAAAGATGACAATAGTAAGTCGA-3' (outside deletion) (Tm=58)
Probe: Fam-5'-ACTACTTTGATTARGCATAAAAGCAAG -3'-Tamra (outside deletion) (Tm=73) *R=A or G
Note: Single Nucleotide Polymorphisms (SNPs) were discovered in both probe segments, and therefore the probes were constructed with degenerate bases at the "K" and "R" positions. The Leftward probe contained a T at the "K" position when blasted against the LVS sequence, but it contained a G when blasted against the SCHU S4 sequence. The Rightward probe contained a G at the "R" position when blasted against the SCHU S4 sequence, but it contained an A when blasted against the LVS sequence.

Panel 2-3b: RT-RD_{Spain} PCR Master-Mix Recipe.

Components:	Source:	Volume (ul)	Final [ ]
2xUMM	Applied Biosystems	12.5	1x
Left-Fwd Primer (4.0uM)	Invitrogen	2.5	400nM
Left-Probe (Vic) (6.0uM)	Applied Biosystems	2.5	600nM
Left-Rev Primer (4.0uM)	Invitrogen	2.5	400nM
Right-Probe (Fam) (6.0uM)	Applied Biosystems	2.5	600nM
Right-Rev Primer (4.0uM)	Invitrogen	2.5	400nM
Platinum Taq (5 U/μl)	Invitrogen	0.11	0.55 U/RXN
Sample Template DNA/Water (Neg)	DNA Collection	1.5	~1.5ng
Total Volume		26.61ul	

Panel 2-3c: RT-RD_{Spain} PCR Cycling Conditions.

Note: *Run FAM & Vic*	Hot Start	PCR			
Hold	Hold	Cycle (45 cycles)			Hold
		Denature	Anneal	Extend	
50°C	95°C	95°C	56°C	61°C	4°C
2 min	10 mins	15 secs	15 sec	1 min	Infinity

Table 2-4. Northern Arizona University *Francisella* DNA Tested by RT-RD_{Spain} PCR.

The table shows the Keim Genetics Laboratory's 319 *Francisella* DNA samples tested by the RT-RD_{Spain} PCR assay. The samples are grouped by numbers of strains according to species and subspecies within geographic regions or countries. The values in the "Type" column refer to the species, P = *F. philomiragia* (if not *F. tularensis*), or subspecies, where A = *tularensis*, B = *holarctica*, N = *novicida*, and M = *mediaasiatica*.

Table 2-4: Northern Arizona University *Francisella* DNA Tested by RT-RD_{Spain} PCR.

Number:	Species	Type	Subspecies	Country
1	*F. tularensis*	A	*tularensis*	Unknown
1	*F. tularensis*	B	*holarctica*	Unknown
2	*F. tularensis*	A	*tularensis*	Canada
1	*F. tularensis*	B	*holarctica*	Canada
1	*F. tularensis*	M	*mediaasiatica*	Central Asia
8	*F. tularensis*	B	*holarctica*	Czech Repub
26	*F. tularensis*	B	*holarctica*	Finland
2	*F. tularensis*	B	*holarctica*	France
7	*F. tularensis*	B	*Holarctica-japonica*	Japan
2	*F. tularensis*	B	*holarctica*	Norway
1	*F. tularensis*	A	*tularensis*	Russia
7	*F. tularensis*	B	*holarctica*	Russia
1	*F. tularensis*	M	*mediaasiatica*	Russia
2	*F. tularensis*	A	*tularensis*	Slovakia
1	*F. tularensis*	B	*holarctica*	Slovakia
2	*F. tularensis*	B	*holarctica*	Spain
127	*F. tularensis*	B	*holarctica*	Sweden
1	*F. philomiragia*	P	Not Applicable	Sweden
5	*F. tularensis*	B	*holarctica*	Ukraine
65	*F. tularensis*	A	*tularensis*	USA
44	*F. tularensis*	B	*holarctica*	USA
1	*F. tularensis*	N	*novicida-like*	USA
4	*F. tularensis*	N	*novicida*	USA
5	*F. philomiragia*	P	Not Applicable	USA
2	*F. tularensis*	M	*mediaasiatica*	USSR
T = 319				

Figure 2-1.

Paired End Sequence Mapping (PESM) Protocol Flowchart.

All steps of the PESM protocol are clearly indicated. Yellow rectangles indicate all steps including generation of the Lambda plaques, direct PCR amplification (DPA) to generate clone amplicons, and sizing of the clone amplicons by gel electrophoresis. White rectangles indicate analytical steps related to sequencing and size quantification, and lastly, to inputting the quantitative size and sequence data in to the PESM Program (or otherwise called, Pipeline).

Figure 2-1: Paired End Sequence Mapping (PESM) Protocol

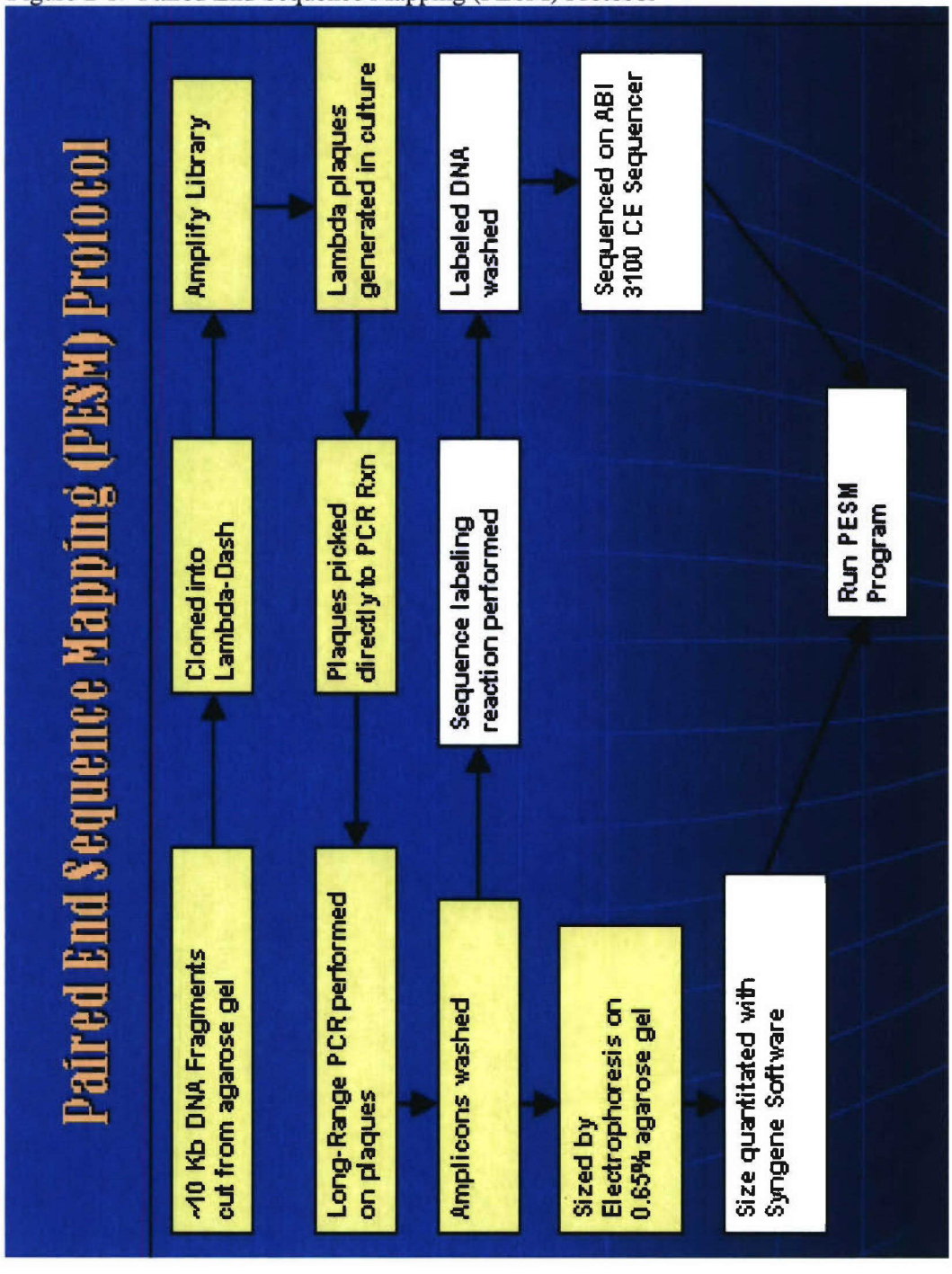


Table 2-5. $CR_{holarctica}$ 1-17 CG-PCR Primer Coordinates, Sequences, and Amplicon Sizes.

For each $CR_{holarctica}$, the primer coordinates, sequences, and predicted amplicons-sizes specific to both LVS (L) and SCHU S4 (S) are listed. The common (C) primer for both the LVS and SCHU S4 coordinates are listed. "SS"=subspecies-specific, "F" =forward, and "R"= reverse.

Table 2-5: $CR_{holarctica}$ 1-17 CG-PCR Primer Coordinates, Sequences, and Amplicon Sizes.

CR#:	Common (C) & Subspec-specific (SS) Primer Coordinates:				Primer Sequences:	
1	C-Primer	SS-LVS	SS-SchuS4	Frag Size (bp):	C-Fwd:	5'-CCTTGATAATCCAAATATGAGTGC-3'
	L-F=2809	R=4546		1737	L-Rev:	5'GTTTTGATTCTATTGACACACCTTG-3'
	S-F=2580		R=4854	2274	S-Rev:	5'-CAAAATATAGCTCCCAGAGATCTAGC-3'
2	C-Primer	SS-LVS	SS-SchuS4	Frag Size (bp):	C-Rev:	5'-CAATGGTTTTATAAACAGCTTCTACG-3'
	L-R=185566	F=184023		1543	L-Fwd:	5'-CTCACAAGGCATTAGATGATATTCG-3'
	S-R=288782		F=285516	3266	S-Fwd:	5'-GTATTTAGGTTCACCAGCTAAAAAGG-3'
3	C-Primer	SS-LVS	SS-SchuS4	Frag Size (bp):	C-FR:	5'-CCCAGTGTCTAATTAGCTTTAGTTGC-3'
	L-FR=301319	R=302900		1581	L-Rev:	5'-GTTGCGGTGTAGGATAATACATATCTC-3'
	S-FR=1532877		F=1530518	2359	S-Fwd:	5'-CAGCTTGCTCGATATTTTCAATAG-3'
4	C-Primer	SS-LVS	SS-SchuS4	Frag Size (bp):	C-Fwd:	5'-CTGAGATATACCTAGAGTCGCAAAAG-3'
	L-F=383837	R=386159		2322	L-Rev:	5'-GCAAACAATAGAAATAGGTAACAAAGC-3'
	S-F=1330804		R=1331699	895	S-Rev:	5'-GGTGCTGGTGAACTTATATATCTGG-3'
5	C-Primer	SS-LVS	SS-SchuS4	Frag Size (bp):	C-Fwd:	5'-GAGAATATTCCATTTGGTGAAACAC-3'
	L-F=442404	R=444118		1714	L-Rev:	5'-CTGTAATGATTGTCCTGCAAATAAC-3'
	S-F=403065		R=405840	2775	S-Rev:	5'-GCTCAACATACTTGATAACCCTATCTC-3'
6	C-Primer	SS-LVS	SS-SchuS4	Frag Size (bp):	C-FR:	5'-GATCAAAGCTCTAAGGTATCAGCTC-3'
	L-FR=832195	R=833676		1481	L-Rev:	5'-CATCAGCACTATTAGGCAAATTCTC-3'
	S-FR=1123552		F=1121191	2361	S-Fwd:	5'-CAATGGTCACTATGATGAACATACC-3'
7	C-Primer	SS-LVS	SS-SchuS4	Frag Size (bp):	C-Fwd:	5'-GCAAGTACTTCTGAAGCTAATCGTATG-3'
	L-F=936709	R=938884		2135	L-Rev:	5'-CTTAATAAATATCCCCAAACCCAAC-3'
	S-F=710085		R=711730	1645	S-Rev:	5'-GGTAAACCCGGTTATTTTACTCAAC-3'
8	C-Primer	SS-LVS	SS-SchuS4	Frag Size (bp):	C-Rev:	5'-CAATGAAGTCCCACTGAGTATATCG-3'
	L-R=1232053	F=1229570		2483	L-Fwd:	5'-CTCTACATACTCGCATAGCTCAGTTG-3'
	S-R=358441		F=356961	1480	S-Fwd:	5'-CAAGCTATGTTTGTAGAAGGTGCTAAG-3'
9	C-Primer	SS-LVS	SS-SchuS4	Frag Size (bp):	C-Rev:	5'-GATGATCGATTGATGCTTAGAAAC-3'
	L-R=1316661	F=1315452		1209	L-Fwd:	5'-GGGTTGCTAGCTATATTGACAAAAC-3'
	S-R=1517541		1515484	2057	S-Fwd:	5'-GCTTTTCTATCACAAACAGAACATAGC-3'
10	C-Primer	SS-LVS	SS-SchuS4	Frag Size (bp):	C-Rev:	5'-GCTGATGATTCTAGCCTTAAAGAAG-3'
	L-R=1388613	F=1386529		2084	L-Fwd:	5'-GCCTGGCATAATTACTGTTTTAGC-3'
	S-R=1332997		F=1331797	1200	S-Fwd:	5'-GCCAATCCTACATTTATAGAACCTG-3'
11	C-Primer	SS-LVS	SS-SchuS4	Frag Size (bp):	C-Rev:	5'-CTCCAACTGCTAATGACTGTACG-3'
	L-R=1432079	F=1430657		1422	L-Fwd:	5'-CAGGAATAAGAGCAACTGCAACTAC-3'
	S-R=738934		F=736815	2119	S-Fwd:	5'-GTGAGTATGGTTGGTATGTATCGTG-3'
12	C-Primer	SS-LVS	SS-SchuS4	Frag Size (bp):	C-FR:	5'-CCGGTTATTTTACTCAACAACTAAATG-3'
	L-FR=1473779	R=1475005		1226	L-Rev:	5'-GTTGGTAATCAGGTTGACATTTTATC-3'
	S-FR=711723		F=709632	2091	S-Fwd:	5'-CTAATGGGTAGAGAGCTGCAAAAG-3'
13	C-Primer	SS-LVS	SS-SchuS4	Frag Size (bp):	C-FR:	5'-GGCTTAGTATTATTGGTGTATGCTTTAG-3'
	L-FR=1572935	R=1574278		1343	L-Rev:	5'-GCAGTAGAAAGAAGTAAGGCTTGG-3'
	S-FR=1650099		F=1647851	2248	S-Fwd:	5'-CTCCACGGATTACAACATTTCC-3'
14	C-Primer	SS-LVS	SS-SchuS4	Frag Size (bp):	C-Fwd	5'-CCCTGCTTTTGCTTCTAGTCC-3'
	L-F=1644147	R=1645543		1396	L-Rev:	5'-CTCGTTCAGCTCTACAACATGC-3'
	S-F=1707477		R=1709692	2215	S-Rev:	5'-CTATATGCTTAGGAGGTGTTTCTGC-3'
15	C-Primer	SS-LVS	SS-SchuS4	Frag Size (bp):	C-FR:	5'-GGCTTATCACCACTGTTCTTGTATC-3'
	L-FR=1691700	R=1693204		1504	L-Rev:	5'-GCAAAACTAAGAGATGGGCTTATG-3'
	S-FR=144633		F=142224	2409	S-Fwd:	5'-GTAGTTTGGGATAAAGAAACAGGTG-3'
16	C-Primer	SS-LVS	SS-SchuS4	Frag Size (bp):	C-Rev:	5'-CTTGTCAGAGTTGGAGTGAAGC-3'
	L-R=1788007	F=1787226		781	L-Fwd:	5'-GTCAAATCACTTTCCTCTCGTTC-3'
	S-R=1801908		F=1799576	2332	S-Fwd:	5'-CCACTACCTCGAATCTTACACAAAG-3'
17	C-Primer	SS-LVS	SS-SchuS4	Frag Size (bp):	C-Rev:	5'-CTATCACCATAGGACTTAATGACTGG-3'
	L-R=1859383	F=1857070		2313	L-Fwd:	5'-GCAAACAGATCCTATAACGTAAGATAGC-3'
	S-R=220265		F=218782	1483	S-Fwd:	5'-GGTTGGCTAGTTGTCAGTATTGG-3'

Table 2-6. Conventional CG-PCR Master-Mix Recipe and Reaction Conditions.

Panel 2-6a shows the the master-mix recipe for each CG-PCR assay. Panel 2-6b shows the thermocycling conditions for each CG-PCR assay.

Panel 2-6a: Conventional CG-PCR Master Mix Recipe.

Components:	Source:	Vol (ul)	Final [ ]
DI H_2O	Invitrogen (Carlsbad, CA)	17.5	1X
10X 50mM Mg 10X Buffer	Idaho Technology (Salt Lake, UT)	2.5	1X
10X dNTPs	Idaho Technology (Salt Lake, UT)	2	160uM each dNTP
Common Primer (25 μM)	Invitrogen (Carlsbad, CA)	0.5	500 nM
Fwd Primer (25 μM)	Invitrogen (Carlsbad, CA)	0.5	500 nM
Rev Primer (25 μM)	Invitrogen (Carlsbad, CA)	0.5	500 nM
Platinum Taq (5 U/μl)	Invitrogen (Carlsbad, CA)	0.5	2.5 U/RXN
Sample Template DNA	DNA Collection	1.5	(1-50 ng/RXN)
Total Volume		25.5	

Panel 2-6b: Conventional CG-PCR Assay Cycling Conditions.

Hot Start	PCR			Final	Final Hold
	Cycle (32 cycles)				
Hold	Denature	Anneal	Extend	Extend	
95°C	95°C	60°C	72°C	72°C	4°C
2 min, 30 sec	30 sec	1 min	1 min	5 min	Infinity

CHAPTER 3:

Phylogeographic Variation In Subpopulations Of *F. tularensis* Subsp. *holartica*: Identification Of A Genomic Marker Of Strains From Northwestern Spain And France

3.0 Abstract

Francisella tularensis, the etiologic agent of tularemia, is comprised of two main subspecies; *F. tularensis* subsp. *tularensis* (Type A) and *F. tularensis* subsp. *holarctica* (Type B). *F. tularensis* subsp. *tularensis* appears to be largely limited to North America and is believed to be more virulent in humans and animals. Comparative genome analyses have identified several regions of difference (RD) in the genomes of *F. tularensis* subsp. *tularensis* strains that distinguish them from the strains of the *holartica* subsp., including a region in the *F. tularensis* Pathogenicity Island (FPI) that encodes virulence-associated genes. To determine whether phylogeographic patterns of genomic variation can be detected in Type A and Type B populations, we have studied strains isolated from the U.S., Europe, and Russia using a shotgun DNA microarray derived from a U.S. Type A strain. The $RD_{tularensis}$ that have previously been identified in independent studies were conserved in strains from both continents. However, searches for geography-specific RD detected a 1.596 Kb deletion in subsp. *holatrtica* isolates from Spain and France. Further analysis of an extended strain set showed that this RD (RD_{Spain}) was limited to subsp. *holartica* isolates from the Iberian peninsula. Phylogenetic analysis of strains carrying RD_{Spain} by multi-locus variable number repeat (VNTR) analysis (MLVA) showed that the strains comprise a highly related set of genotypes, implying that they constitute (one or) two clonal populations sharing recent common ancestry. In conjunction with epidemiological evidence, our data collectively support the conclusion that strains sharing the RD_{Spain} represent a recent introduction and/or clonal expansion of a subsp. *holartica* subclone in the Iberian Peninsula.

3.1 Introduction

F. tularensis is a Gram-negative, facultative intracellular pathogen originally isolated from ground squirrels in 1911 during a plague investigation in Tulare County, CA [1]. The organism is believed to affect more animal species than any other known zoonotic pathogen [2, 3]. Historically, the organism has been weaponized and is considered a significant biowarfare agent, especially due to its extremely low infectious dose when acquired through the inhalation route in humans [4].

Although the species *F. tularensis* is comprised of four subspecies—subsp. *tularensis*, subsp. *holarctica*, subsp. *mediaasiatica*, and subsp. *novicida*—epidemiological evidence suggests only the *tularensis* and *holartica* subspecies appear to be clinically significant to humans [100]. Even between the *holartica* and *tularensis* subspecies, there appears to be a hierarchy of infectivity with the *tularensis* subspecies being considered more infectious for humans [13, 14]. A simple hypothesis to explain differences in infectivity is that the subsp. *tularensis* and subsp. *holartica* populations have significantly different genome content with respect to virulence associated genes. Studies in independent laboratories using DNA microarray-based CGH to measure genome diversity in *F. tularensis* populations of different geographic origin [90, 91] identified eight different RD distinguishing all subsp. *tularensis* from subsp. *holartica* strains tested. These tularensis-specific genomic regions include $RD_{tularensis}6$, a segment of the FPI encompassing the *pdpD* gene recently shown to facilitate virulence in F. *tularensis* subsp. *tularensis* [58]. Therefore, at least one explanation for differences in infectivity for humans may be the content of FPI in the two different subspecies.

In addition to differences in virulence characteristics, the *F. tularensis* subsp. *tularensis* and subsp. *holartica* populations show differences in geographic distribution, with the *tularensis* subsp. being largely limited to North America. To determine if phylogeographic patterns of genome variation can be detected among the *holartica* populations from the U.S. and Europe, we have now analyzed a set of *F. tularensis* strains from the U.S., continental Europe, and the former

Soviet Union by comparative genome hybridization (CGH) using our previously defined *F. tularensis* subsp. *tularensis* shotgun DNA microarray [91]. Here we show that the previously identified $RD_{tularensis}$ are conserved among *holartica* strains from both continents and have identified a new RD, termed RD_{Spain}, that is found in strains from Spain and France. Analysis of strains derived from global locations shows that RD_{Spain} is confined to strains from Spain and France (and a single isolated recovered from Sweden). Phylogenetic analysis of strains carrying the RD_{Spain} further showed that the strains comprise one to two clonal populations. Collectively, our data show that phylogeographic variation can be detected in the genomes of *F. tularensis* subsp. *holartica* populations and imply that recent clonal expansion has accompanied epidemic spread of a subpopulation of the *holartica* subspecies on the Iberian Peninsula.

3.2 Results

(Materials and Methods are presented in Chapter 2, sects. 2.1 and 2.2)

3.2.1 $RD_{tularensis}$ Are Conserved in Globally Derived *Francisella* Strains.

Our previous study of genome variation among *F. tularensis* subsp. *tularensis* and subsp. *holartica* populations from the U.S. identified 13 $RD_{tularensis}$ that were unique to subsp. *tularensis* strains. Only a subset of these RD was observed in the studies of Broekhujisen *et al.* [90], which used a different array and a strain set largely derived from Europe. To test whether the $RD_{tularensis}$ unique to each study were due to representation on the respective arrays or whether it is due to phylogeographic variation in the genomes of the *F. tularensis* populations on the North American and European continents, we first used CGH to compare strains of *F. tularensis* subsp. *holartica* from the US and from Europe. The strain set comprised forty-eight different strains In addition to forty-two subsp. *holarctica* strains obtained from Spain, the strain set also included three strains from Czech Republic (Tu-28, Tu-29, and Tu-35), one from the Russian Federation (Tu-42)

and a F. tularensis subsp. *novicida* (Tu-43) strain from the U.S. These strains were examined alongside a set of 16 strains that had previously been tested on this same array [91].

When the data from hybridizations of U.S. and European strains was composited and sorted for subspecies-specific alterations, a total of 48 probes were identified that hybridized exclusively to *F. tularensis* subsp. *tularensis* strains and not to any of the subsp. *holartica* strains (see figure 3-1). DNA sequencing and contig analysis of these probes revealed that the probes were derived from ($RD_{tularensis}1$ - $RD_{tularensis}13$) that had previously been identified in our study of strains exclusively from the U.S. [91]. This finding confirms that these RD are indeed conserved broadly among subsp. *holartica* strains from both continents and the finding is consistent with the $RD_{tularensis}$ occurring early during divergence of populations of the *tularensis* subsp. and the *holartica* subsp from their common ancestor.

3.2.2 Phylogeographic Variation Within Subsp. *holartica* Populations From Europe.

Despite the conservation of the previously identified $RD_{tularensis}$, sorting of the data between taxa of subsp. *holartica* strains from the US compared to the large set of strains from Spain identified five probes that detected polymorphisms specific to these strains (see figure 3-2). DNA sequence analysis revealed that all five probes formed a 1.8 Kb contig (see figure 3-3), corresponding to the FTT1006-FT1008 coding regions from the SCHU S4 genome sequence [48]. To precisely map the endpoints of the deletion, PCR primers were designed within the FTT1005c and FTT1009 coding regions abutting the 1.8 Kb contig of array probes and PCR was performed on *F. tularensis* subsp. *tularensis* and *F. tularensis* subsp. *holartica* strains. As shown in Figure 3-4, each of the *F. tularensis* subsp. *holartica* strains that showed deviation from the reference strain (LVS) in the CGH gave rise to the same size PCR product of 1.38 Kb, whereas all other strains produced a 2.97 Kb amplicon in common with the SCHU S4 strain. DNA sequence analysis of the shorter 1.38 Kb fragment from the Spanish-outbreak strain, Tu-19,

showed that the deletion extends from the FTT1006-FTT1008 coding regions, as predicted from the microarray data.

The deletion begins in the central region of the FTT1006 (403 AA, hypothetical membrane protein) coding region at amino acid $Leucine_{160}$ (SCHU S4 Coord: 1019329) and extends into the FTT1008c (54 AA, hypothetical protein) coding region, ending at codon $Leucine_{23}$ (SCHU S4 Coord: 1020924). The protein encoded by FTT1006 shares similarity with several different proteins, including transporter proteins, and *flg*J-like amidasases. FTT1007 and FTT1008 encode hypothetical proteins. It was also noted that the two regions flanking RD_{Spain} contain sequences with homology to IS element-associated genes (a transposase and resolvase – see figure 3-5) as was observed for the two previous CGH studies [90, 91]. To further confirm the distribution of the deletion, PCR primers were designed and used in a nested PCR strategy (Chapter 2 - Materials and Methods) to examine DNA from ninety isolates (compared to ninety-six that were tested by the conventional RD_{Spain} PCR assay), including a broad temporal range (ranging from 1920 to 2004) and spatial distribution of subsps. *tularensis*, *holarctica*, and *novicida*, as well as a *F. philomiragia* strain, the live vaccine strain (LVS) of subsp. *holartica*, and SCHU S4 as a subsp. *tularensis* reference control.

As shown in Table 3-1, all subsp. *holartica* strains and all subsp. *tularensis* strains, except all 42 from Spain and all four from France, were negative for the deletion, producing the predicted-sized band of 2.97 Kb. Strains from Spain and France all produced the same sized 1.38 Kb band predicted from the RD_{Spain} allele. The four strains from France included two previously contained in the AFIP strain collection and two more obtained from France during this investigation. The two former from the AFIP collection are named (after villages in France from/near where they were isolated) Chateneaux (near Midwestern France) and St. Germaine (from Southeastern France near Carmaux), or AFIP3 and AFIP, respectively. Of the two new isolates, "FR-LR" was isolated from Lorraine (near Nancy which is in Northeastern France), and

"FR-SS" was isolated from a region near Langres (South and West of Nancy) from a human patient with a case of *F. tularensis* bacteremia [101].

3.2.3 The RD_{spain} Is Limited To Isolates From The Iberian Peninsula.

To further explore distribution of the RD_{Spain}, a large DNA panel (n=319) of *F. tularensis* subsp. *tularensis* (n=71), subsp. *holartica* (n=233), subsp. *novicida* (n=5), subsp. *mediaasiatica* (n=4), and *F. philomiragia* (n=6) strains isolated from Europe, Asia, and North America was tested for the presence of the RD_{Spain} deletion. The strain set comprised 319 strains from the Keim Genetics Laboratory (see Table 3-2 for results and geographic distribution for all 319 strains) which have previously been characterized by MLVA and represent a broad genetic diversity of *F. tularensis* populations [2]. DNA from each strain was tested using the nested RT-PCR assay designed to detect the RD_{Spain} (Chapter 2 - Materials and Methods). Figure 3-6 shows a representative RD_{Spain} RT-PCR results-output screen after testing these samples. Of the 319 strains, 314 carried an intact region whereas the remaining five had RT-PCR products consistent with the RD_{Spain} deletion. Of the five isolates carrying the RD_{Spain} two, F0020 and F0295, were isolated from the Voges and Chateauroux regions of France, respectively; two were isolated from Spain -- F0284 from an unspecified region, and F0326 isolated from Madrid; and unexpectedly, one was isolated from Uppsala Sweden (FF0228). Previous study of the genetic relationship of these strains shows that they comprise a distinct cluster among the subsp. *holartica* strains examined [2]. As we demonstrate below, these strains are highly related to those isolated from Spain and likely represent an epidemic spread of a highly related population.

3.2.4 Strains Carrying RD_{spain} Comprise Two Unique Clonal Populations.

Many of the strains from Spain that were used in our study have previously been shown to share a close genetic relationship by AFLP and other genotyping methods [32, 33]. In order to place them relative to the recently inferred phylogeny of global *Francsiella* populations, strains

carrying the RD_{Spain} were genotyped by MLVA and compared to MLVA data generated previously [2, 76] following normalization of the data to account for system (instrumentation, reagents, etc.) differences between the two laboratories.

Strains carrying the RD_{Spain} comprised two primary genotypes which differ from one another by only two MLVA alleles at the M4 and M22 loci (see Table 3-3). All RD_{Spain}-positive strains from the Keim Genetics Laboratory (otherwise known as the NAU Laboratory, or NAU) carried a 443 bp and a 254 bp allele at these two loci (yellow highlights), respectively; whereas all AFIP RD_{Spain}-positive strains had a 440 bp and a 249 bp allele at the same two loci (green highlights), respectively.

Additional molecular subtyping within the RD_{Spain} subpopulation was evident due to the combinations of poly-allelism at the M3, M5, M6, M10, M23, and M24 loci. Based on these differences, the AFIP subset was grouped into eight total molecular subtypes (alternate term for genotypes). In addition, four separate molecular subtypes were observed within the NAU Laboratory's RD_{Spain}-positive subset on the basis of poly-allelism at the M3, M6, M8, and M24 loci (shown in Table 3-3 by different colored highlighting).

To confirm the finding of two primary genotypes rather than just one as we originally hypothesized, we sent to the NAU Laboratory DNA from eight of our AFIP RD_{Spain}-positive strains (FR-SS, FR-LR, Tu-3, Tu-5, Tu-9, Tu-31, Tu-38, and Tu-41), with each one representing one of our eight identified molecular subtypes. These DNA samples were tested along with the NAU RD_{Spain}-positive strain DNAs by direct-comparative selective (DCS) MLVA at the M4 and M22 loci. Additionally, seven of the DNAs sent to NAU (all but FR-SS) were tested for the allele(s) responsible for differentiating each respective genotype from the others. The results of this DCS-MLVA essentially confirmed the original findings demonstrating two unique primary genotypes; but two of the strains were resolved into previously observed genotypes based on two alleles each, therefore reducing the number of genotypes in the AFIP collection from eight to six. Whereas Ft-31 previously had allele sizes at M5 and M10 of 208 bp and 617 bp, respectively,

DCS-MLVA resulted in allele sizes of 192 bp and 361 bp, respectively. Also, whereas Ft-41 previously had allele sizes at M23 and M24 of 435 bp and 480 bp, respectively, DCS-MLVA resulted in allele sizes of 412 bp and 465 bp, respectively. The results of the original NAU strains remain as originally obtained.

The only isolate carrying RD_{Spain} which was not originally isolated from the Peninsula was F0228, a strain isolated from a human patient in Upsala, Sweden in 2000 (resulting from a mosquito bite-derived lesion) during part of a large epidemic in that country [P. Keim, personal communication]. However, its MLVA genotype is quite distinct from strains comprising that outbreak [2] and differs as well among all other RD_{Spain}-positive strains in this study by having alleles not shared by any others at M3, M6, and M8. Collectively, these data suggest that the origin of F0228 may be distinct from strains comprising either of the two respective outbreaks.

With respect to all RD_{Spain}-negative (or wild-type) strains listed in Table 3-3, all RD_{Spain}-positive strains demonstrated a unique M24 allele of 465 bp with the exception of F0020 which had an allele size of 461 bp.

3.3 Discussion

The potential for phylogeographic variation in *F. tularensis* populations seems reasonable given the differences in global distribution of the *F. tularensis* subsp. *tularensis* and *F. tularensis* subsp. *holartica* populations. MLVA and other genotyping methods also provide some support for geographic differentiation in populations, but the populations are not entirely resolvable into geographic clusters [2, 26, 102]. Our study using DNA microarray analysis has shown that phylogeographic variation can be detected at the whole genome level without sequencing, and the variation is concordant with phylogenetic analysis. In conjunction with epidemiological data, we believe the variation that was observed in our study is a consequence of recent epidemic spread of a highly-related clonal population that has apparently undergone additional divergence.

Therefore, it is possible that geographic clustering observed in various genotypic studies of *F. tularensis* may indeed represent expansion of populations of regional clones.

It is noteworthy that this specific polymorphism has not previously been characterized until now as demonstrated by recent reviews of the literature [10, 17]. While the authors maintain the long-held assertion that subsp. *holarctica* strains are overall less virulent to humans than those of subsp. *tularensis*, it is interesting that the RD_{Spain} polymorphism is associated with the first recorded tularemia outbreak in NW Spain, which occurred between 1997 and 1998, resulting in over 500 human cases [103], and which required hospitalization of at least 142, nearly 22.5% of whom experienced initial therapeutic failures as previously mentioned in chapter 1. In addition, the fact that the French-bacterimia strain contains the RD_{Spain} polymorphism is also noteworthy. Collectively these epidemiological data suggest that RD_{Spain}-positive strains may represent a hypervirulent subsp. *holarctica* subpopulation among humans. It should not be surprising that the loss of genes could result in increased virulence since it has been shown that *Yersinia pestis*, the etiological cause of plague and descendant of the less pathogenic *Y. psuedotubercolosis,* is the product of the loss of 317 genes from the latter opposed to the gain of only 31 new genes in the former [104]. If this is the case for RD_{Spain}-*holarctica* strains, however, remains speculative for now, and validation of such a hypothesis would require functional genomic studies involving knockout mutants and animal models. Besides the limited gene homologues identified in the deleted DNA segment, that segment may also encode *cis* elements such as promoter or enhancer sites associated with regulation of virulence factors. The discovery of the RD_{Spain} polymorphism in a recent subspec. *holarctica* isolate in Sweden where the overwhelming majority of isolates are wild-type is not well understood at this time. Whereas the patient denied any travel to either Spain or France, it is possible that this isolate is a sentinel indicator of emergence from Spain or France, perhaps via exported lagomorphs or domestic pets.

Regardless of whether this RD_{Spain}-positive subpopulation emerged once or twice, resolution of this matter is likely not as important as understanding its global, including regional,

geographic distribution (the current geographic distribution of subsp. *holarctica* RD_{Spain} variant strains is represented in a map in figure 3-7). Therefore, isolates from other countries surrounding France, Spain, and Sweden not yet represented in our combined collections (i.e., Portugal, Italy, Germany, Belgium, United Kingdom, Denmark, and the Netherlands) should be tested to provide an accurate distribution of the RD_{Spain} -*holarctica* subspecies. This should help to ensure DoD Force Health Protection and public safety within endemic areas.

Table 3-1.

Master AFIP-UNMC RD_{Spain} Strain-DNA Collection .

For each sample from the collection, the table shows the strain- or DNA-identification (ID) number, original subspecies determination, spatial and temporal information, host or vector ID, and results for the conventional-RD_{Spain},RT-RD_{Spain}, and RD1 PCR assays. DNA samples carrying the RD_{Spain} polymorphism were considered mutants (mut) and marked with red rectangles, with respect to wild-type (WT) which were marked with green rectangles. All "mut" DNA samples were exclusively from Spain and France, and all were from the *holarctica* subspecies as verified by RD1 PCR. The single *F. philomiragia* DNA sample was negative for RD_{Spain} PCR as shown by yellow highlighting.

Table 3-1: Master AFIP-UNMC RD_{Spain} Strain-DNA Collection Sheet-1

Strain ID# used in Study	Species or subspecies	Geographic origin, year of isolation	Host or Vector	Convent. RD_{Spain} PCR	RT-RD_{Spain} PCR	Conv. Subsp. RD1 PCR
Tu-30	*tularensis*	Utah, 1920	Human	WT	WT	Type-A
SchuS4	*tularensis*	Strain SchuS4, Ohio, 1941	Human	WT	WT	Type-A
A88R160	*tularensis*	Strain A88R160, USA	Rabbit	WT	WT	Type-A
LVS	*holarctica*	Russia, 1961	Unknown	WT	WT	Type-B
Tu-1	*holarctica*	Valladolid, Spain, 1997	Hare	Mut	Mut	Type-B
Tu-2	*holarctica*	Valladolid, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-3	*holarctica*	Valladolid, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-4	*holarctica*	León, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-5	*holarctica*	Palencia, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-6	*holarctica*	León, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-7	*holarctica*	León, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-8	*holarctica*	Palencia, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-9	*holarctica*	Palencia, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-10	*holarctica*	Valladolid, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-11	*holarctica*	Valladolid, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-12	*holarctica*	Zamora, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-13	*holarctica*	Zamora, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-14	*holarctica*	Palencia, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-15	*holarctica*	Valladolid, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-16	*holarctica*	Valladolid, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-17	*holarctica*	Segovia, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-18	*holarctica*	Palencia, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-19	*holarctica*	Palencia, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-20	*holarctica*	Palencia, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-21	*holarctica*	Palencia, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-22	*holarctica*	Valladolid, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-28	*holarctica*	Strain 130, Czech Republic	Hare	WT	WT	Type-B
Tu-29	*holarctica*	Strain 2713, Czech Republic	Hare	WT	WT	Type-B
Tu-35	*holarctica*	Strain T-1/59, Czech Republic	Hare	WT	WT	Type-B
Tu-36	*holarctica*	Palencia, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-37	*holarctica*	Valladolid, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-38	*holarctica*	Soria, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-39	*holarctica*	Zamora, Spain, 1999	Hare	Mut	Mut	Type-B
Tu-40	*holarctica*	Zamora, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-44	*holarctica*	Ávila, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-45	*holarctica*	Valladolid, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-47	*holarctica*	León, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-48	*holarctica*	Palencia, Spain, 1998	Hare	Mut	Mut	Type-B
Tu-31	*holarctica*	Valladolid, Spain, 1998	Human	Mut	Mut	Type-B
Tu-32	*holarctica*	Valladolid, Spain, 1998	Human	Mut	Mut	Type-B
Tu-33	*holarctica*	Valladolid, Spain, 1998	Human	Mut	Mut	Type-B
Tu-34	*holarctica*	Palencia, Spain, 1998	Human	Mut	Mut	Type-B
Tu-24	*holarctica*	León, Spain, 1998	Human	Mut	Mut	Type-B
Tu-25	*holarctica*	Zamora, Spain, 1998	Human	Mut	Mut	Type-B
Tu-26	*holarctica*	Zamora, Spain, 1998	Human	Mut	Mut	Type-B
Tu-27	*holarctica*	Zamora, Spain, 1998	Human	Mut	Mut	Type-B
AFIP3	*holarctica*	Chateneaux, France	Human	Mut	Mut	Type-B
AFIP4	*holarctica*	St. Germaine, France	Human	Mut	Mut	Type-B

Table 3-1: Master AFIP-UNMC RD_{Spain} Strain-DNA Collection Sheet-2

Strain ID# used in Study	Species or subspecies	Geographic origin, year of isolation	Host or Vector	Convent. RD_{Spain} PCR	RT-RD_{Spain} PCR	Conv. Subsp. RD1 PCR
Tu-23	*holarctica*	Zamora, Spain, 1998	Vole	Mut	Mut	Type-B
Tu-41	*holarctica*	Zamora, Spain, 1998	Tick	Mut	Mut	Type-B
Tu-46	*holarctica*	Valladolid, Spain, 1998	Tick	Mut	Mut	Type-B
Tu-42	*holarctica*	Strain 503, Russia	Tick	WT	WT	Type-B
F.t.n15482	*novicida*	ATCC 15482, Utah, 1950	Water	WT	WT	*novicida*
Tu-43	*novicida*	Texas, 1991	Human	WT	WT	*novicida*
D2005067002	*novicida*	Unknown	Unknown	WT	WT	*novicida*
F.ph 25015	*F. philomiragia*	ATCC 25015, Utah, 1959	Muskrat	Neg	Neg	Neg
99A-2628	*holarctica*	Strain 99A-2628, California	Human	WT	WT	Type-B
89A-7092	*holarctica*	Strain 89A-7092, California	Squir. Monkey	WT	WT	Type-B
Japanese	*holarctica*- japan	Japan	Unknown	WT	WT	*japonica*
Austrian	*holarctica*	Vienna	Unknown	WT	WT	Type-B
FSG-1	*holarctica*	Georgia, FSU, 1987	Tick	WT	WT	Type-B
FSG-2	*holarctica*	Georgia, FSU, 1980	Tick	WT	WT	Type-B
FSG-3	*holarctica*	Georgia, FSU, 1977	Bird	WT	WT	Type-B
FSG-4	*holarctica*	Georgia, FSU, 1997	Tick	WT	WT	Type-B
FSG-5	*holarctica*	Georgia, FSU, 1974	Com. Shrew	WT	WT	Type-B
FSG-6	*holarctica*	Georgia, FSU, 2002	Vole	WT	WT	Type-B
FSG-7	*holarctica*	Georgia, FSU, 2002	Tick	WT	WT	Type-B
FSG-8	*holarctica*	Georgia, FSU, 1997	Vole	WT	WT	Type-B
FSG-9	*holarctica*	Georgia, FSU, 1990	Vole	WT	WT	Type-B
FSG-10	*holarctica*	Georgia, FSU, 1956	Gerbil	WT	WT	Type-B
88R52	*tularensis*	Strain 88R52, USA, 1988	Rabbit	WT	WT	Type-A
88R144	*tularensis*	Strain 88R144, USA, 1988	Rabbit	WT	WT	Type-A
AK-1133496	*tularensis*	Fairbanks, Alaska, 2003	Arctic Hare #1	WT	WT	Type-A
AK-1100558	*tularensis*	North Pole, Alaska, 2004	Arctic Hare #2	WT	WT	Type-A
AK-1100559	*tularensis*	North Pole, Alaska, 2004	Arctic Hare #2	WT	WT	Type-A
FR-LR	*holarctica*	Lorraine, France, 1993	Human	Mut	Mut	Type-B
FR-SS	*holarctica*	Near Langres, France, 2000	Human	Mut	Mut	Type-B
UNMC061598	*tularensis*	NE Ref Strain, Nebraska	Human	WT	WT	Type-A
UNL091902	*tularensis*	Nebraska, USA	Human	WT	WT	Type-A
WY-WSVL01	*holarctica*	Wyoming, USA	Bovine	WT	WT	Type-B
WY-9868529	*holarctica*	Wyoming, USA	Guinea Pig	WT	WT	Type-B
WY-00W4114	*tularensis*	Wyoming, USA	Prairie Dog	WT	WT	Type-A
WY-96194280	*holarctica*	Wyoming, USA	Rabbit	WT	WT	Type-B
WY-WSVL02	*tularensis*	Wyoming, USA	Human	WT	WT	Type-A
OK-00101504	*tularensis*	Oklahoma, USA	Feline	WT	WT	Type-A
OK-98041035	*tularensis*	Oklahoma, USA	Feline	WT	WT	Type-A
MS-304	*holarctica*	Missouri, USA	Human	WT	WT	Type-B
NC-54558-01	*tularensis*	North Carolina, USA	Feline	WT	WT	Type-A
NC-52797-99	*tularensis*	North Carolina, USA	Rabbit	WT	WT	Type-A
NC-54559-01	*tularensis*	North Carolina, USA	Feline	WT	WT	Type-A
CDC NE 031457	*tularensis*	Lincoln, NE, USA, 2003	Human	WT	Not Tested	Type-A
UNL 072704	*tularensis*	Lincoln, NE, USA, 2004	Rabbit	WT	Not Tested	Type-A
Rabbit N. Cd.	*tularensis*	North Carolina, USA	Rabbit	WT	Not Tested	Type-A
MO MSI349	*tularensis*	Missouri, USA	Human	WT	Not Tested	Type-A
AFIOH Feline	*tularensis*	Oklahoma, USA	Feline	WT	Not Tested	Type-A
MO N. Cd. Human	*holarctica*	Missouri, USA	Human	WT	Not Tested	Type-B

Figure 3-1.

Comparative Genomic Hybridization (CGH) Microarray Analysis.

The figure shows a CGH profile of a Spanish outbreak sample (Tu#1) compared to a U.S. *subsp.* tularensis (Tu#30) CGH profile. Red spots correlate with RD-associated addresses where sequence from the test-DNA is absent or different relative to the reference-DNA present on the chip.

Figure 3-1: Comparative Genomic Hybridization (CGH) Microarray Analysis.

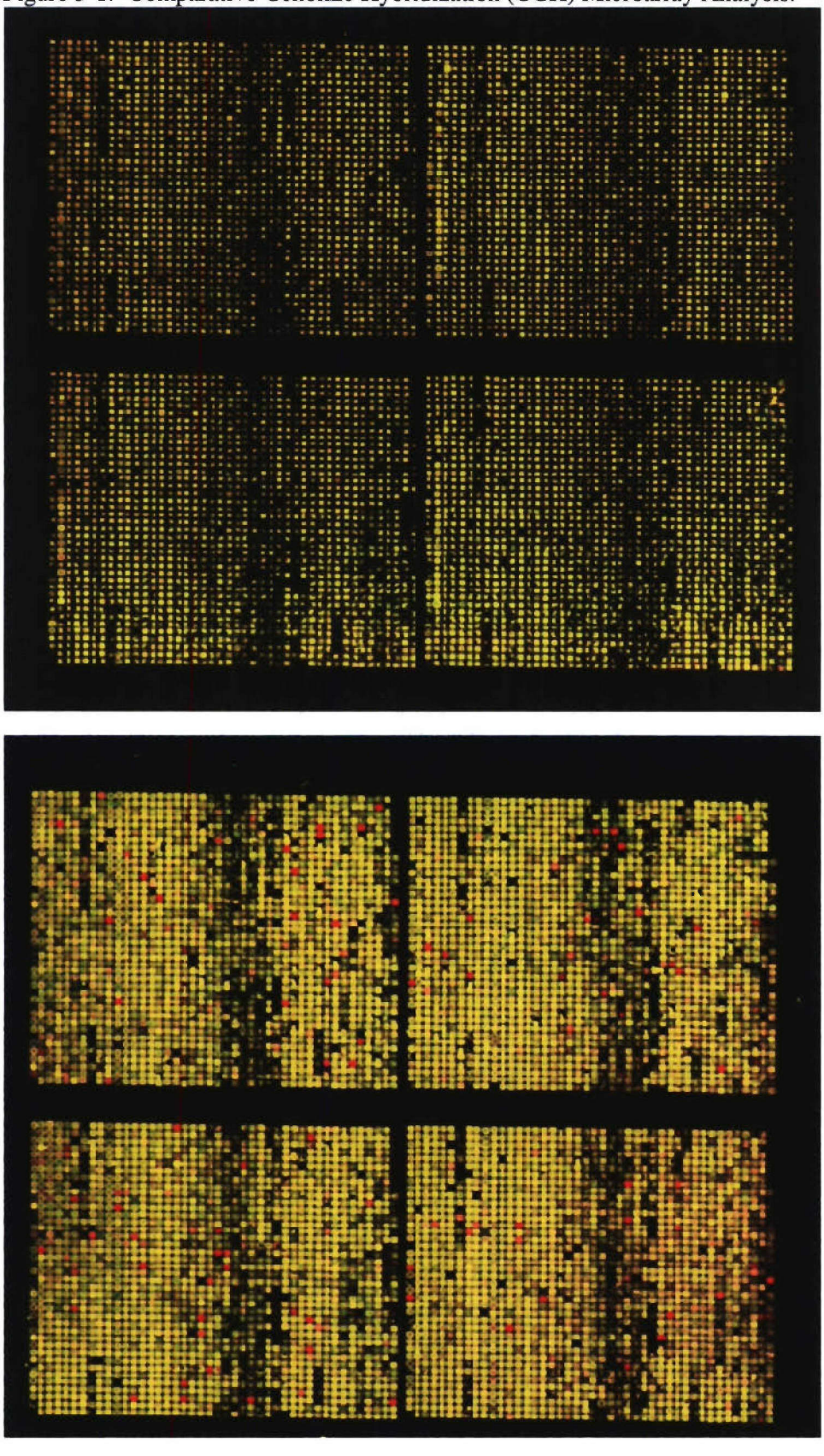


Figure 3-2.

Microarray Data Shows Subsp. *holarctica* Phylogeographic Variation.

Shown in the figure is an UPMGA dendogram showing variation in subsp. *holartica* DNA samples between Spanish outbreak populations (blue) and US and other European (red) populations. All "TU" strains from the Spanish outbreak are lacking five array addresses (red chads) compared with non-Spanish outbreak strains shown. The two lower "TU" samples containing the five array addresses are from the Czech Republic.

Figure 3-2: Microarray data shows subsp. *holarctica* phylogeographic variation.

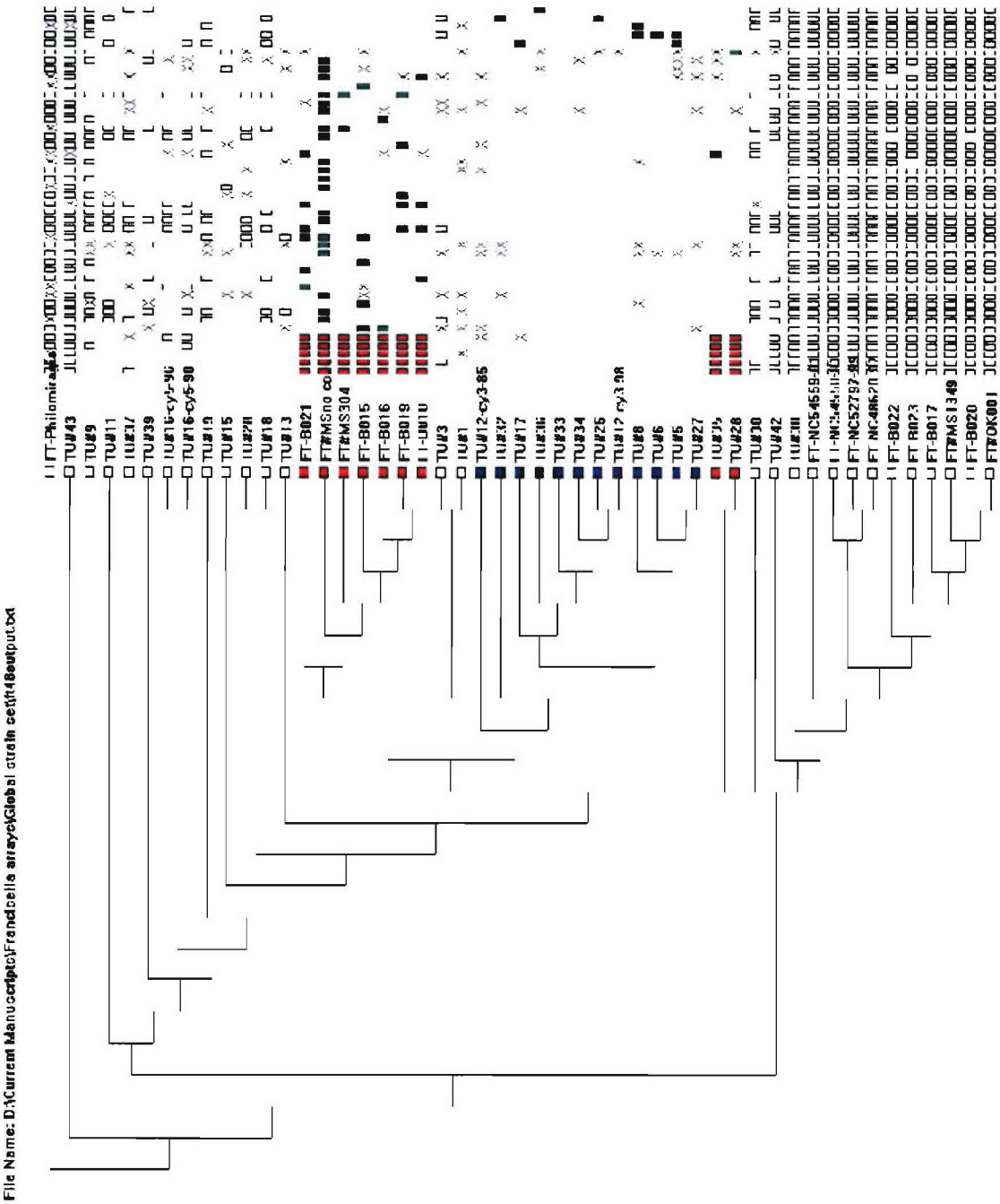


Figure 3-3.

Contig Assembly of Microarray Addresses.

The figure shows the single 1.8 kb contig (large green line) formed from alignment of sequences from the five corresponding reference subsp. *tularensis*-specific array addresses.

Figure 3-3: Contig Assembly of Microarray Addresses.

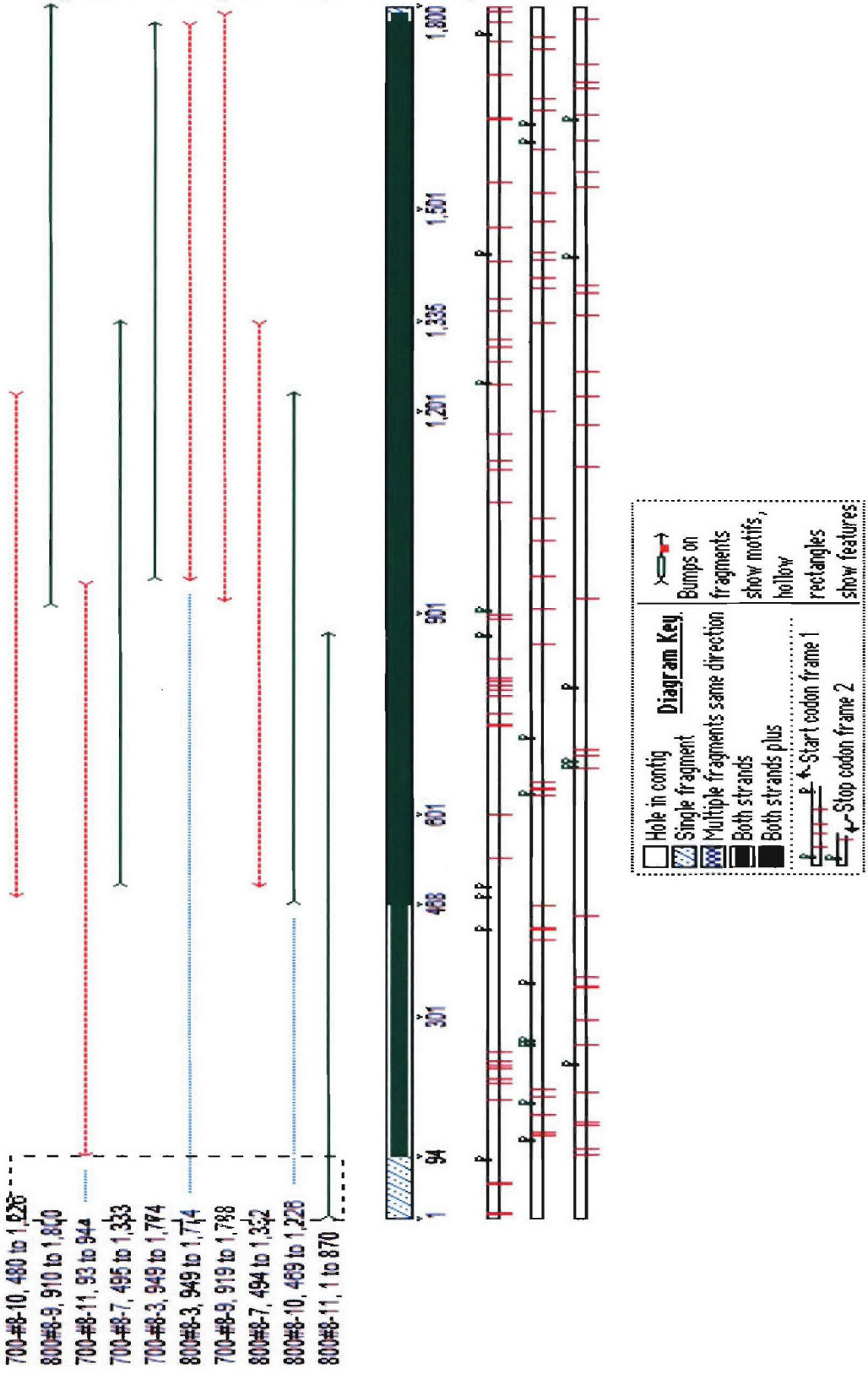


Figure 3-4.

Panel 3-4a: Conventional RD_{Spain} PCR Assay.

The panel shows examples of PCR bands corresponding to the RD_{Spain}-deletion (Tu-4), and wild-type (WT) for all others. Bv-A=Biovar-A (Type-A=subspec. *tularensis*), whereas Bv-B=Biovar-B (Type-B=subsp. *holarctica*). The coded samples are as follows: UNL Ref=NE-UNL091902; BO19=WY-9868529; BO23=OK-98041035; and NC54-8-01= NC-54558-01. All others are as listed in Table 2-1.

Panel 3-4b: Conventional Subspecies and RD_{Spain} PCR Assays.

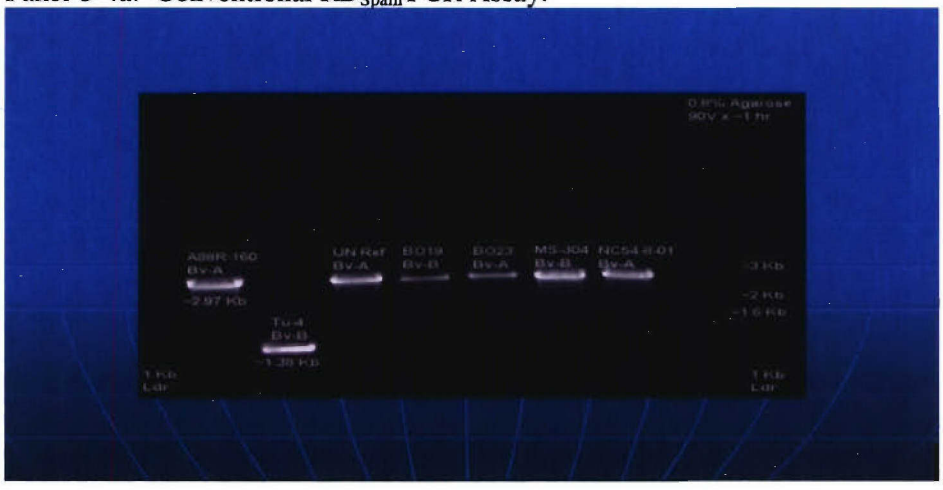
The same DNA samples were loaded for both the upper and lower rows. The upper row shows examples of c34-5 subspecies PCR bands corresponding to subsp. *tularensis* (T-A=Type A, i.e., SCHU S4) and subsp. *holarctica* (T-B=Type B, i.e., LVS). These results were reproduced with RD1 subsp. PCR (data not shown). The bottom row shows wild-type PCR bands, such as those for both SCHU S4 and LVS, and RD_{Spain}-deletion bands, such as for Tu-23. Names of abbreviated samples are as follows: AK-496=Alaska-1133496; AK-558=Alaska-1100558; AK-559=Alaska-1100559.

Panel 3-4b: Conventional RD1 Subspecies PCR Assay.

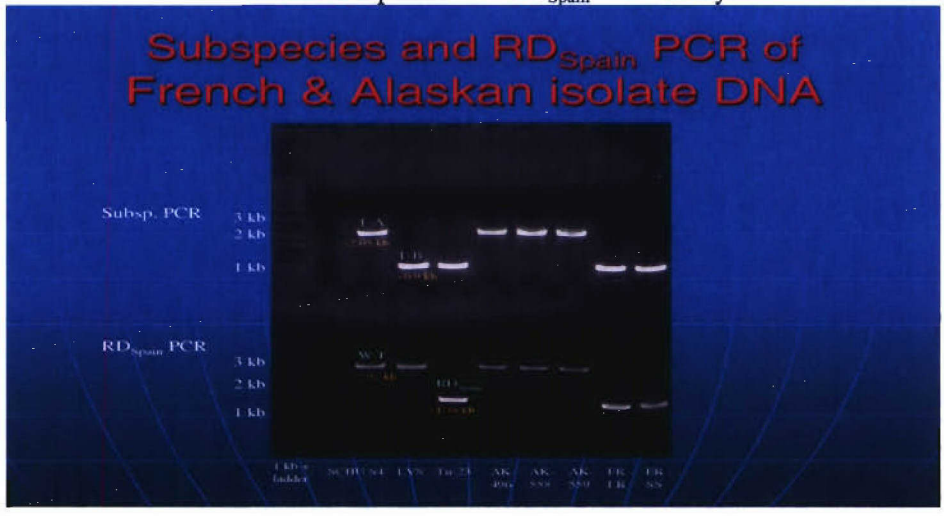
The panel shows examples of RD1 subspecies PCR bands corresponding to subsp. *holarctica* (T-B) and SCHU S4 (T-A). Also shown are bands corresponding with NE Ref (UNMC 061598), A88R-160, Grousse (an AFIP subsp. *holarctica* strain of unknown origin), AFIP-Vienna, AFIP-Jap, Tu-23, FSG-10, *F .philomiragia* 25015, and *F. tularensis* subsp. *novicida*. The no template control (NTC) lane and the *F. philomiragia* lane are negative, and all other samples produced the expected-sized amplicons.

Figure 3-4

Panel-3-4a: Conventional RD_{Spain} PCR Assay.



Panel 3-4b: Conventional Subspecies and RD_{Spain} PCR Assays.



Panel 3-4c: Conventional RD1 Subspecies PCR Assay.

Figure 3-5.

ORFs Showing Hypothetical Proteins.

The RD_{Spain}-deletion region (as shown here from LVS) contains only a few hypothetical proteins including one with sequence homology to a sodium-dependant transport pump. The deletion region (highlighted in yellow) is flanked by a transposase and resolvase.

Figure 3-5. ORFs showing hypothetical proteins.

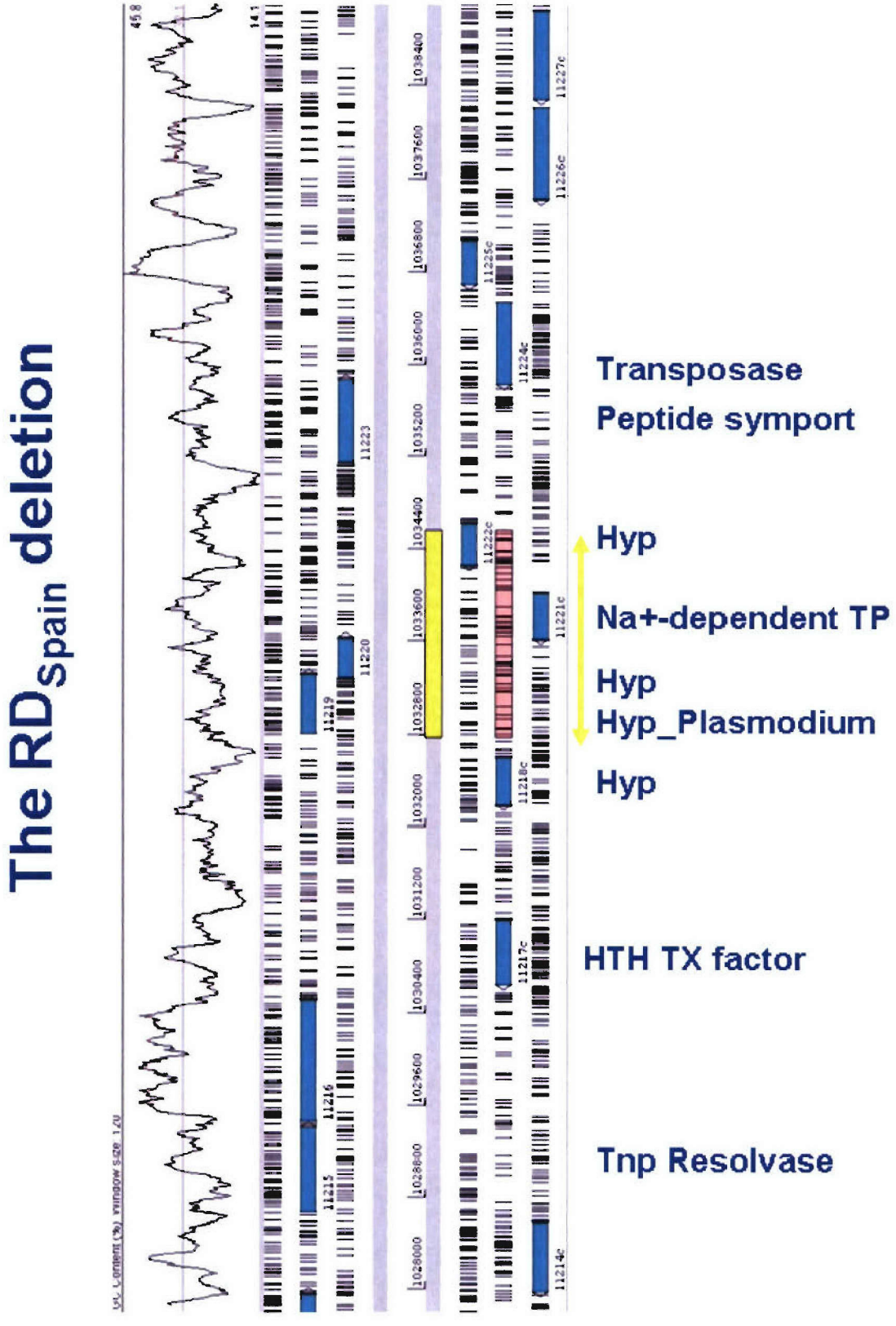


Figure 3-6.

Schematic of RT-RD_{Spain} PCR Design and Allelic Discrimination Output Screen.

Shown in the upper-right panel of the figure is the line-diagram of the RT-RD_{Spain} PCR assay, where the leftward-forward primer serves to prime both possible mutually exclusive allelic amplicons, and therefore allowing production of only one of the two possible fluorescent reactions on the ABI 7900 HT instrument. The lower-right panel shows representative output "Allelic Discrimination" screen from the ABI 7900 HT, where the RD_{Spain}-deletion positive samples are shown in the upper left of the grid, the NTCs are shown in the lower left, and the wild-type samples are shown in the lower right of the grid.

Figure 3-6: Schematic of RT-RD_{Spain} PCR Design and Allelic Discrimination Output Screen.

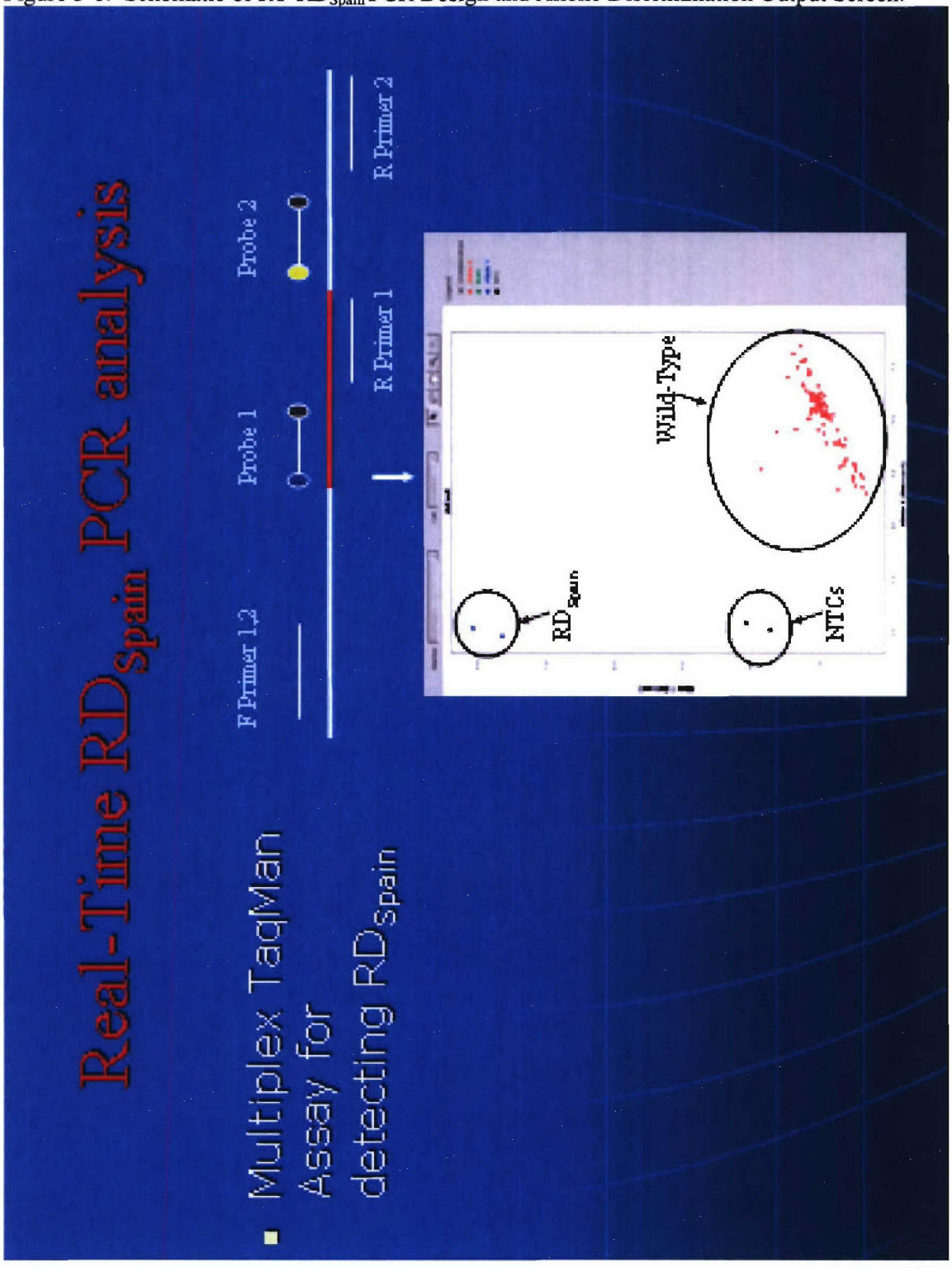


Table 3-2.

Results from RT-RD_{Spain}-PCR Testing of the NAU Global *Francisella* Panel.

As shown in the table, out of a total of 319 strain DNAs tested, only five were positive for the RD_{Spain}-deletion (results shown in red and highlighted with fluorescent-green). Two of the positive were from France, two were from Spain, and unexpectedly, one was from Sweden. The "Type" column denotes the species, P=*F. philomiragia*, or the *F. tularensis* subspecies, where A=*tularensis*, B=holarctica, M=*mediaasiatica*, and N=*novicida*.

Table 3-2: Results from RT-RD_{Spain}-PCR Testing of the NAU Global *Francisella* Panel.

Number:	Species	Type	Subspecies	Country	Results
1	*F. tularensis*	A	*tularensis*	Unknown	wild type
1	*F. tularensis*	B	*holarctica*	Unknown	wild type
2	*F. tularensis*	A	*tularensis*	Canada	wild type
1	*F. tularensis*	B	*holarctica*	Canada	wild type
1	*F. tularensis*	M	*mediaasiatica*	Central Asia	wild type
8	*F. tularensis*	B	*holarctica*	Czech Repub	wild type
26	*F. tularensis*	B	*holarctica*	Finland	wild type
2	*F. tularensis*	B	*holarctica*	France	Positive
7	*F. tularensis*	B	*Holarctica-japonica*	Japan	wild type
2	*F. tularensis*	B	*holarctica*	Norway	wild type
1	*F. tularensis*	A	*tularensis*	Russia	wild type
7	*F. tularensis*	B	*holarctica*	Russia	wild type
1	*F. tularensis*	M	*mediaasiatica*	Russia	wild type
2	*F. tularensis*	A	*tularensis*	Slovakia	wild type
1	*F. tularensis*	B	*holarctica*	Slovakia	wild type
2	*F. tularensis*	B	*holarctica*	Spain	Positive
126	*F. tularensis*	B	*holarctica*	Sweden	wild type
1	*F. tularensis*	B	*holarctica*	Sweden	Positive
1	*F. philomiragia*	P	*Not Applicable*	Sweden	wild type
5	*F. tularensis*	B	*holarctica*	Ukraine	wild type
65	*F. tularensis*	A	*tularensis*	USA	wild type
44	*F. tularensis*	B	*holarctica*	USA	wild type
1	*F. tularensis*	N	*novicida-like*	USA	wild type
4	*F. tularensis*	N	*novicida*	USA	wild type
5	*F. philomiragia*	P	*Not Applicable*	USA	wild type
2	*F. tularensis*	M	*mediaasiatica*	USSR	wild type
T = 319					

Table 3-3.

NAU and AFIP *F. tularensis* MLVA Results with ABI 377-Normalized Scores.

Two main genotypes of RD_{Spain}-positive strains are shown based on M4 and M22 loci. Molecular subtypes are shown as color coded strain DNA numbers/names grouped together as determined by the hypervariable M3 locus and occasionally other differential loci. All RD_{Spain}-positive strains except F0020 had an M24 allele size of 465 bp whereas none of the other subsps. *tularensis* or *holarctica* strains tested had such an allele size. The subspecies are shown by "Type", with B=*holarctica* and A=*tularensis.*

Table 3.3: NAU and AFIP *F. tularensis* MLVA Results with ABI 377-Normalized Scores (sheet-1)

NAU DNA #	RD_{Spain}	Type	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24	M25
FD020 (France)	Pos	B	310	451	297	443	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	254	412	461	156
FD228 (Sweden)	Pos	B	310	451	396	443	192	353	175	286	216	361	392	345	313	443	227	211	345	361	285	254	396	254	412	465	156
FD284 (Spain)	Pos	B	310	451	297	443	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	254	412	465	156
FD295 (France)	Pos	B	310	451	297	443	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	254	412	465	156
FD320 (Spain)	Pos	B	310	451	369	443	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	254	412	465	156
AFIP DNA #	RD_{Spain}	Type	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24	M25
Tu-3	Pos	B	310	451	315	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-5	Pos	B	310	451	324	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-38	Pos	B	310	451	297	440	192	332	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-6	Pos	B	310	451	288	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-9	Pos	B	310	451	288	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-1	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-4	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-7	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-12	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-15	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-16	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-17	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-18	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-20	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-21	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-22	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-24	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-25	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-31	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-34	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-36	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-44	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-48	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
FR Syl-S	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
FR Chateneaux	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
FR St. Germaine	Pos	B	310	451	297	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156

Table 3-3: Sheet 1.

Table 3-3: Sheet 2.

Table 3.3: NAU and AFIP *F. tularensis* MLVA Results with ABI 377-Normalized Scores (sheet-2)

AFIP DNA #	RD_{Egyptn}	Type	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24	M25
Tu-2	Pos	B	310	451	306	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-6	Pos	B	310	451	306	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-10	Pos	B	310	451	306	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-11	Pos	B	310	451	306	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-13	Pos	B	310	451	306	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-14	Pos	B	310	451	306	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-19	Pos	B	310	451	306	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-23	Pos	B	310	451	306	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-26	Pos	B	310	451	306	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-32	Pos	B	310	451	306	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-33	Pos	B	310	451	306	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-37	Pos	B	310	451	306	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-39	Pos	B	310	451	306	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-41	Pos	B	310	451	306	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-45	Pos	B	310	451	306	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-46	Pos	B	310	451	306	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
Tu-47	Pos	B	310	451	306	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
FR Lau-R	Pos	B	310	451	306	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	465	156
MS-304	Neg	B	310	451	360	443	192	353	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	480	156
Tu-28	Neg	B	310	451	297	440	192	353	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	480	156
Tu-29	Neg	B	310	451	324	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	480	156
Tu-35	Neg	B	310	451	297	440	192	353	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	480	156
Tu-42	Neg	B	310	451	387	440	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	480	156
WY 986829	Neg	B	310	451	360	443	192	332	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	480	156
WY Bovine	Neg	B	310	451	360	443	192	332	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	480	156
WY 96194280	Neg	B	310	451	405	443	192	332	175	254	216	361	392	345	313	443	227	211	345	361	285	254	396	249	412	480	156
MO N.Cd. Human	Neg	B	310	451	350	443	192	311	175	254	216	361	392	345	313	443	227	211	345	361	285	254	417	249	412	480	156
Tu-30	Neg	A	310	463	492	436	192	311	207	286	248	361	392	345	325	443	222	220	351	370	316	254	403	243	435	461	156
AFIOH Feline	Neg	A	310	451	414	436	224	332	191	270	295	489	392	345	325	443	222	220	345	370	316	254	396	243	435	461	156
UNLD72704	Neg	A	310	463	315	436	208	332	207	286	280	489	392	345	325	443	222	220	351	370	316	254	403	243	435	461	156
UNLD91902	Neg	A	310	463	333	436	208	332	191	286	295	568	392	345	325	443	222	220	351	370	316	254	403	243	435	461	156
CDC NE03 1457	Neg	A	310	463	333	436	208	332	191	286	295	568	392	345	325	443	222	220	351	370	316	254	403	243	435	461	156

Table 3.3: NAU and AFIP *F. tularensis* MLVA Results with ABI 377-Normalized Scores (sheet-3)

AFIP DNA #	RD_{Spain}	Type	M 1	M 2	M 3	M 4	M 5	M 6	M 7	M 8	M 9	M 10	M 11	M 12	M 13	M 14	M 15	M 16	M 17	M 18	M 19	M 20	M 21	M 22	M 23	M 24	M 25
NC 52797-99	Neg	A	310	505	405	436	192	311	191	254	264	361	392	345	325	443	222	220	351	370	316	254	403	243	412	461	156
NC 5455801Feline	Neg	A	310	559	342	436	192	311	191	254	280	361	392	345	325	443	222	220	351	370	316	254	396	243	412	461	156
SCHUS4	Neg	A	310	463	432	436	208	311	207	286	248	617	392	345	325	443	222	220	351	370	316	254	403	243	435	461	156
AK 133496	Neg	A	310	529	572	436	192	311	175	254	216	537	392	345	325	443	222	220	351	370	316	254	403	243	412	461	156
AK 100558	Neg	A	310	493	477	436	192	311	175	254	216	537	392	345	325	443	222	220	351	370	316	254	403	243	412	461	156
AK 100559	Neg	A	310	463	477	436	192	311	175	254	216	537	392	345	325	443	222	220	351	370	316	254	403	243	412	461	156
UNMC-061598	Neg	A	310	469	405	436	192	332	175	302	328	457	392	345	325	443	222	220	351	370	316	254	410	243	435	461	156
WY Human-NC-2	Neg	A	310	487	252	436	192	311	175	254	216	345	392	345	313	443	227	211	345	361	316	491	403	249	412	461	156
WY00W4114	Neg	A	310	559	342	440	192	332	175	254	216	345	392	345	313	443	227	211	345	361	316	477	396	249	412	461	156
OK 100101504	Neg	A	310	463	414	436	224	332	191	270	295	489	392	345	325	443	222	220	351	370	316	254	396	243	435	461	156
MOM51349	Neg	A	310	463	342	436	240	332	191	270	264	568	392	345	325	443	222	220	351	370	316	254	410	243	435	461	156
Rab.N.Cd.AFIOH	Neg	A	310	601	369	436	192	311	191	270	264	457	392	345	325	443	222	220	351	370	316	254	417	243	412	461	156

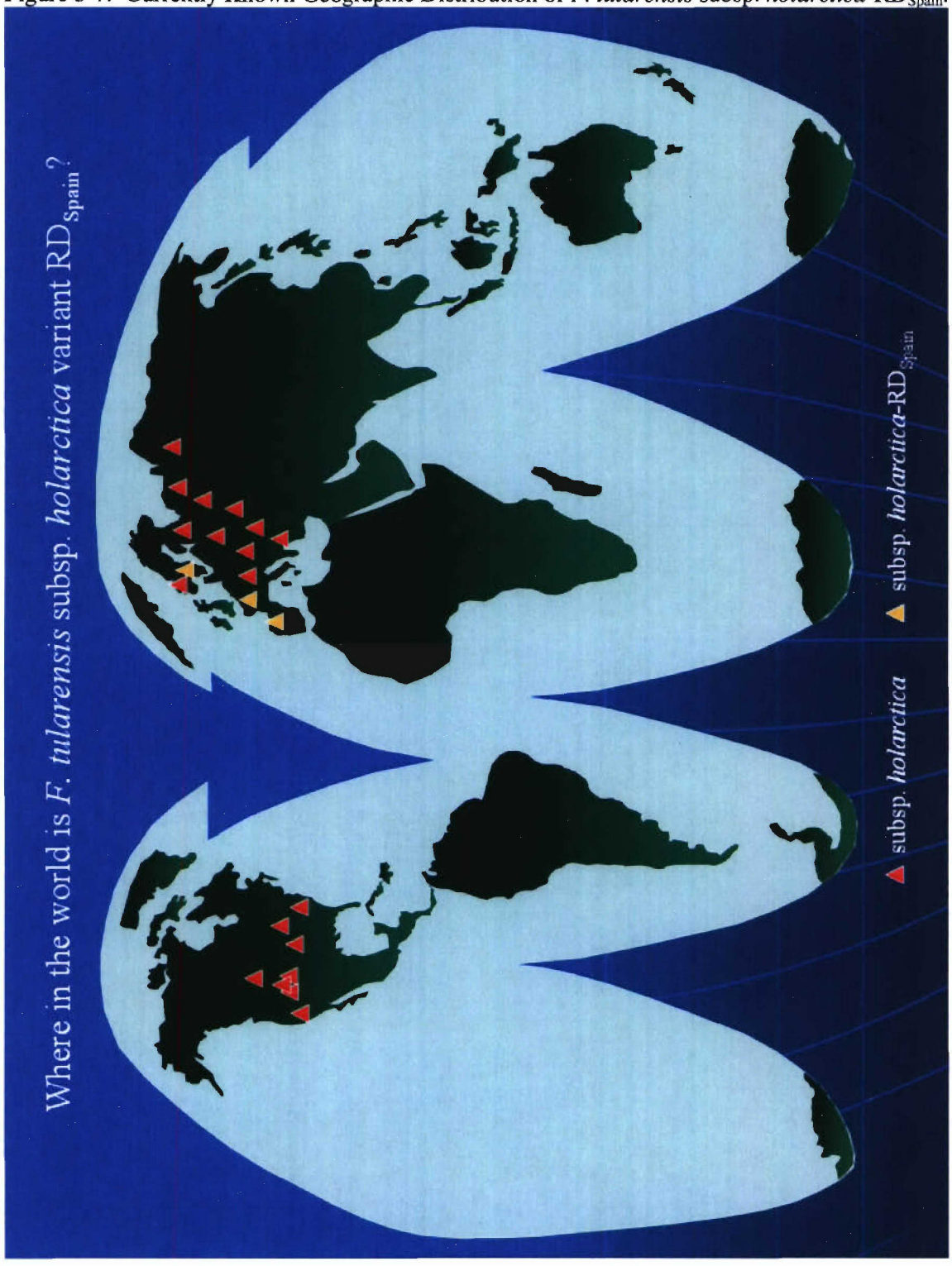
Table 3-3: Sheet 3.

Figure 3-7.

Currently Known Geographic Distribution of *F. tularensis* subsp. *holarctica*-RD$_{Spain}$.

Shown in the map are red triangles representing geographic areas correlating with subsp. *holarctica*, and yellow triangles representing known locations of RD$_{Spain}$ variants of subsp. *holarctica*.

Figure 3-7: Currently Known Geographic Distribution of *F. tularensis* subsp. *holarctica*-RD_{Spain}.



CHAPTER 4:

Paired-End Sequence Mapping Detects Extensive Genomic Rearrangement And Translocation During Divergence Of *Francisella tularensis* Subspecies *tularensis* And *Francisella tularensis* Subspecies *holarctica* Populations.

4.0 Abstract

Comparative genome analyses of the *Francisella tularensis* subspecies *tularensis* and subspecies *holarctica* populations show that genome content is highly conserved and only a relatively small number of genes within the ~1.9 Mb *F. tularensis* subsp. *tularensis* genome are absent in other *F. tularensis* subspecies. To catalogue differences in genome organization that could contribute to the unique virulence characteristics and geographic distributions of the *tularensis* and *holarctica* subspecies, we have used Paired-End Sequence Mapping (PESM) to identify regions of the genome that are non-contiguous between these two subspecies. Using PESM, the physical distances between paired-end sequencing reads from a library of a wildtype reference *F. tularensis* subsp. *holartica* strain were compared to the predicted lengths between the reads based on map coordinates of the reads from the subsp. *tularensis* strain SCHU S4 and subsp. *holarctica* strain LVS genome sequences. A total of 17 different continuous regions were identified in the subsp. *holartica* genome ($CR_{holartica}$) that are non-contiguous in the subsp. *tularensis* genome. At least six of the seventeen different $CR_{holarctica}$ are positioned as adjacent pairs in the subspecies *tularensis* genome sequence but are translocated in *holarctica*, implying that arrangements of the $CR_{holarctica}$ segments are ancestral in the *tularensis* subspecies and derived in *holarctica*. Using nested-PCR assays, the conservation of the events was further assessed by testing 88 additional *tularensis* and *holarctica* subspecies isolates. The PCR results showed that the arrangements of the $CR_{holarctica}$ are highly conserved, particularly in the *holarctica* subspecies, consistent with the hypothesis that subsp. *holarctica* populations have recently experienced a periodic selection event or they have emerged from a recent clonal expansion. Two unique

tularensis-like strains were also observed to share some $CR_{holarctica}$ with the *holarctica* subspecies and others with the *tularensis* subspecies, implying that these strains may represent a new taxonomic unit.

4.1 Introduction

Francisella tularensis is a non-motile, Gram-negative coccobacillus originally isolated from ground squirrels in 1911 during a plague investigation in Tulare County, CA [1]. The geographic distribution of the organism spans the entire Northern Hemisphere, with only a very recent isolated recovery of the organism occurring in the Southern Hemisphere [22, 23]. The organism is a facultative intracellular pathogen and is believed to affect more animal species than any other known zoonotic pathogen [2, 3]. It has been isolated from as many as 250 species of wildlife (reviewed by Oysten, Sjostedt, *et. al*) [10] including various birds, amphibians, fish and many mammalian species. The organism can also be found in invertebrates species, including arthropod vectors such as mosquitoes and ticks (reviewed by Petersen and Schriefer) [17]. Human infection occurs most often through direct exposure to infected animals or by bites from infected arthropod vectors. Recently, terrestrial and aquatic life cycles have been described for *F. tularensis* [18, 19]; and protozoa, such as *Acanthamoeba castellanii*, may also serve as a host for maintenance of *F. tularensis* in the aquatic cycle [21].

The species *F. tularensis* is comprised of four recognized subspecies: Subsps. *tularensis* (Type A), *holarctica* (Type B), *novicida*, and *mediaasiatica*, the two former of which are considered clinically significant in humans [13, 14] and by far have been the most studied. *F. tularensis* subsp. *tularensis* is believed to be more virulent in humans than *F. tularensis* subsp. *holarctica* based on epidemiological data and its higher infectivity in animals. *F. tularensis* subsp. *tularensis* and *F. tularensis* subsp. *holarctica* also show striking geographic differences in their distribution, with both the *tularensis* and *holarctica* subsp. being found in North America but only the *holarctica* subsp. being found in Europe and Asia [17]. Populations of subsp.

mediaasiatica may be even more geographically limited since, as its name suggests, this subspecies has only been isolated from the Asian subcontinent. The *novicida* subspecies has been found primarily in the U.S. but was recently detected in Australia [13, 25].

Despite the unique geographic and virulence characteristics, known genetic and phenotypic differences distinguishing the *tularensis* and *holarctica* subpopulations seem to be more limited. Biochemically, the two subspecies have classically been differentiated primarily on the basis of glycerol fermentation, production of citrulline ureidase, and erythromycin resistance [13]. High resolution genotyping methods such as pulsed field gel electrophoresis (PFGE) [33], restriction-fragment length polymorphism (RFLP) [26], Amplified Fragment Length Polymorphisms (AFLP) [33], and Multi-Locus Variable Number Tandem Repeat analysis (MLVA) [2, 76, 78], also distinguish the subspecies genotypically and show that they are divergent, but clonally related.

Given the unique geographical and virulence characteristics, there is tremendous interest in understanding the genetic basis for these characteristics. Recent comparative genome hybridization studies identified limited differences in genome content between the two subspecies, but did include deletion in the *pdpD* region which is associated with virulence [90, 91]. Comparative genome sequencing efforts are also underway and promise to provide detailed information with regard to specific strains. To provide a more complete catalogue of the genomic events which arose early during divergence of the subspecies (true supsbepcies-specific genomic differences as opposed to strain-level differences), we have applied Paired End Sequence Mapping (PESM) to identify candidate regions of genomic difference and further used Comparative Genome PCR (CG-PCR) on a large set of strains to identify regions of genomic difference that are conserved across multiple isolates. PESM was originally developed as a method to identify genomic islands of *Shigella dysenteriae* [105]. The PESM strategy measures the physical distance between paired-end reads from a clone library, specifically searching for clones whose physical distance is incongruent with the predicted distance based on available

genome sequences. In this application, we constructed a library from an *F. tularensis* subsp. *holarctica* strain and compared physical distances with the *F. tularensis* SHU S4 genome sequence. Cloned segments with incongruent lengths compared to the map position were further distinguished as strain-specific versus potentially subspecies-specific by comparison to the *F. tularensis* subsp. *holarctica* strain LVS genome sequence. In instances where the length difference was conserved in the reference strain and the LVS *holarctica* strain, the segments were further tested among a panel of *holarctica* and *tularensis* strains to confirm that the genome difference was broadly conserved across the subspecies. Using this strategy, we identified seventeen regions in the genome that are continuous in 66 of 67 subspecies *holarctica* strains examined, but which are discontinuous in *tularensis* strains. These regions, termed $CR_{holarctica}$, have arisen through extensive insertion/deletion, translocation, and rearrangement events and their conservation among *holarctica* strains of distinct temporal and geographic origin implies that this subspecies has likely been through a recent periodic selection event.

4.2 PESM Results

(Materials & Methods are presented in Chapter 2, sects. 2.1 and 2.3)

4.2.1 Paired-End Sequencing.

A total of 752 plaques were picked and subjected to DPA with 551 of the DPA yielding amplicons >8 Kb in length that were of sufficient quality and quantity for size determination and DNA sequence analysis (DPA success rate of 73.3%). One entire sequence run out of seventeen, however, failed to meet these requirements, most likely attributable to improper sequence reaction master-mix setup, and was therefore discarded. So, from the remaining sixteen sequence and sizing runs, the mean amplicon/insert size from the 551 successful DPA reactions was 14,174 bp, which corresponds to approximately 7.8 Mb of coverage, or an estimated 4.1 X coverage of the 1.89 Mb *F. tularensis* subsp. *tularensis* genome [48]. These statistics are summarized in

figure 4-1. A representative agarose-gel of the sixteen clone-sizing electrophoresis experiments is shown in figure 4-2. Representative Excel screen-capture images from the PESMP and subsequent sorting of this data into $CR_{holarctica}$ segments are shown in figure 4-3.

4.2.2 Mapping of Paired-End Sequence Reads.

Of the 551 clones with quality paired-end reads, 66 clones had physical lengths that were not congruent with distance between the paired-end reads relative to the SCHU S4 genome, but were congruent with distances predicted from the LVS genome. These clones were further considered as candidates for subspecies-specific genomic events. Alignment of the sequences from the paired-end reads of candidate clones grouped the 66 cloned segments into 17 different contiguous regions ($CR_{holarctica}$) which align with the *F. tularensis* subsp *holarctica* strain LVS genome but are non-contiguous or otherwise altered in the *F. tularensis* subsp. *tularensis* SCHU S4 genome sequence. Plotting of the number of $CR_{holarctica}$ identified versus the total number of clones sequenced showed that the number of new $CR_{holarctica}$ began to decrease sharply after the clones from plate #10 were sequenced (see figure 4-1). Of the last ~250 clones sequenced, only three new $CR_{holarctica}$ were identified, suggesting that the library was nearly saturated. The resultant sequence coordinates from the LVS WGS as well as the corresponding coordinates from SCHU S4 are shown in table 4-1.

4.2.3 Comparative Genome PCR (CG-PCR) Confirmation Of $CR_{holarctica}$.

Conservation of the $CR_{holarctica}$ in the MS-304 reference strain—which was isolated in 2002, and is temporally and geographically distinct from the LVS strain (strain pedigree information is shown in table 2-1) isolated in 1941—leads to the simple hypothesis that these CR likely arose early during divergence of the *tularensis* and *holarctica* subspecies and therefore should be conserved across most *holarctica* strains. To confirm this, CG-PCR assays were developed for each $CR_{holarctica}$ using nested primer sets at the junctions of the CR (see table 2-5 for

list of CG-PCR coordinates and primers). The different CG-PCR reactions for all 17 $CR_{holarctica}$ were then run on a panel of DNA samples from SCHU S4 and LVS, and also from 19 different subsp. *tularensis* strains, 67 subsp. *holarctica* strains, 3 subsp. *novicida* strains, and a single strain each of subsp. *holarctica*-japan (also referred to as subsp. *holarctica-japonica*, or subsp. *japonica*) and *F. philomiragia*. Figure 4-4 shows a representative agarose-gel following CG-PCR at one of the seventeen $CR_{holarctica}$ loci. Table 4-2 shows the resultant genotypes (with only the differences summarized) obtained after testing the entire *Francisella* panel with all 17 CG-PCR assays. The results for all *Francisella* strain DNAs tested by all 17 CG-PCR panels are shown in figure 4-5. The colors correspond to different-sized amplicons produced from each CG-PCR reaction. Overall, the subsp. *holarctica* strains produced homogenous results across all 17 CR, with 66 of the strains (98.5%) producing the expected amplicon based on the LVS genome sequence. The only deviation occurred in *F. tularensis* subsp. *holarctica* strain Tu-42, which produced a subsp. *tularensis* A-type band. Thus, excluding this one exception, the $CR_{holarctica}$ identified through the PESM pipeline are indeed highly conserved.

As demonstrated in figure 4-5, unlike the 67 subsp. *holarctica* strains, the 19 subsp. *tularensis* strains displayed significantly more heterogeneity in the CG-PCR assays. At least four different subgroups of the *tularensis* subspecies can be resolved. All share the RD1 region in common, implying that they are taxonomically true subsp. *tularensis* derivatives. SCHU S4 and nine other subsp. *tularensis* strains comprise one subgroup and have identical genome organization, producing A-type bands (red squares), across all 17 CR. A second subgroup is represented by A88R160, 88R52, 88R144, AK-1133496, AK-1100558, AK-1100559, with each of these strains sharing an amplicon from the CR10 nested PCR reaction that was unique in size (denoted by orange squares). All six of these strains were isolated from rabbits or hares, with the three AK strains derived from Alaska in 2003 and 2004 and the three others isolated from the contiguous United States. The ATCC 6223 strain was likewise shown to have an identical CG-PCR genotype as the previous six strains, but it was originally isolated from a human patient in

1920 and has since lost its virulence [2]. A third subgroup is represented by the single strain OK-98041035 which matches the SCHU S4 subgroup except that it failed to produce a CR14 PCR amplicon (denoted by a yellow square). The fourth subgroup comprises a very unique set of two isolates, strains WY-00W4114 and WY-WSVL02. These strains both produced an A-type band at CR3, CR4, CR5, CR6, CR8, CR11, CR15, and CR17 (red squares), were negative at CR1 and CR13 (yellow squares), and produced B-type bands at CR7, CR9, CR12, and CR14 (green squares). Unlike any other strains, they also produced unique bands at CR16 (blue-grey) and CR10 (blue-grey, different in size from orange). These two strains were also distinguishable from one another in that WY-00W4114 produced a B-type band at CR2 (green) whereas WY-WSVL02 was negative (yellow). These two strains were also only slightly capable of fermenting glycerol [91]. Collectively, the genetic and biochemical data strongly suggest these two strains represent a new taxanomic unit. If indeed this is a new taxon, then the population is likely to be virulent since one of the isolates was obtained from a human clinical sample [91].

As would be expected, the *F. tularensis*, subsp. *novicida*, subsp. *holarctica*-japan, and *F. philomiragia* strains showed heterogenous CG-PCR results. *F. philomiragia* was negative (yellow) across each of the 17 CR. The three subsp. *novicida* strains were negative (yellow) for CR1, CR4, CR14, and CR16; and they produced a unique size amplicon from (blue-grey) CR3, CR5, CR7, CR8, CR9, CR10, CR12, CR13, and CR17. All three strains produced a "B"- type allele (green) across CR6 and CR15 and they all produced an "A"-type allele (red) across CR11. CR2 differentiated between the Tu-43 strain, which produced a unique amplicon while the other two *novicida* strains (from ATCC and USAMRIID) were both negative. Consistent with its classification as a separate subspecies [2], the single subsp. *holarctica*-japan strain was also distinct from all other strains in this study; it produced a "B"-type allele (green) across CR1, CR2, CR3, CR5, CR6, CR7, CR9, CR10, CR11, CR12, CR15, CR16, a unique allele (orange) across CR13, an "A"-type allele (red) across CR17, and was negative (yellow) across CR4, CR8, and CR14.

4.2.4 Fine-Structure Analysis of $CR_{holarctica}$.

Fine-structure mapping and annotation of the $CR_{holarctica}$ was next conducted by alignment of the $CR_{holarctica}$ contigs from strain LVS genome sequence with the SCHU S4 genome sequence. The corresponding locations of the aligned regions for all 17 $CR_{holarctica}$ are mapped onto one circular map of the SCHU S4 genome as shown in figures 4-6. In addition, individual line-drawings for all 17 CR are shown in figure 4-9. Figure 4-7 shows individual circular alignments of six of the $CR_{holarctica}$ segments onto the circular SCHU S4 map. The combined DNA represented by the 17 CRs corresponds to nearly 230 genes/pseudogenes and over 30 IS-elements, mainly a combination of ISftu1 and ISftu2 elements. Most of the of the rearrangements and translocations are juxtaposed to IS elements, suggesting that many of the events were likely mediated by these elements, resulting in remarkably large changes in the location of specific genome segments between SCHU S4 and LVS, but with little effect on the corresponding content of the transposed/translocated regions. With regard to content of the CR, it should be noted that nearly all of the genes within the $CR_{holarctica}$ are indeed present in both the subsp. *holarctica* LVS and subsp. *tularensis* SCHU S4 genome, albeit at unique positions.

As shown in figure 4-6, the distribution of the CR1-CR17 segments around the LVS genome and the relative positions of the corresponding regions in the SCHU S4 genome have some remarkable characteristics. First, the CR in the LVS genome show some positional bias, with thirteen of the seventeen $CR_{holarctica}$ being present in roughly one-half of the genome (the region between 8 o'clock and 2 o'clock extending from 1.3 Mb to 0.3 Mb). Secondly, there are three notable instances where segments from different $CR_{holarctica}$ in LVS are juxtaposed to one another in the SCHU S4 genome. Specifically, segments from CR1 and CR16 are adjacent in the SCHU S4 genome as are segments from CR13 and CR15, and CR4 and CR10, and all three events are illustrated in figure 4-8. The juxtapositioning of these segments in subsp. *tularensis* suggests that their organization in the tularensis subspecies was the ancestral state while their

organization in *holarctica* is a derived state. As shown in figure 4-8, panel c, this notion is further supported by the finding that both CR13 and CR15 contain genes that are involved in glycerol fermentation and it seems likely that their ancestral condition would have been functionally clustered.

4.2.5 Genes Affected By Rearrangements.

Further bioinformatics analysis of the junction of the juxtaposed CR4-CR10 segment revealed the complete deletion of a gene of unknown function (FTT1308c) from LVS as compared with its intact presence in the SCHU S4 genome. In addition, nine genes were found which are disrupted as a consequence of the rearrangements or translocation of genome segments in the CR. The intact versions of these genes in the SCHU S4 genome encode proteins with significant similarity to oligopeptide transporters (*oppD* and *oppF*), a ribosome modification gene (*rimK*), an acetyltransferase (FTT0177c), and genes of unknown function (corresponding to FTT0898c, FTT1122, FTT0921, and FTT1311). Figure 4-8c illustrates the region of CR13 in SCHU S4 containing the intact *oppD* and *oppF* genes which are truncated in the LVS genome. The *aceF* gene, encoding the E2 of pyruvate dehydrogenase lies near the junction of CR3 (along with *aceE* and *lpd*) and carries a 300 base in-frame deletion. A fine-structural genome comparison of the entire $CR_{holarctica}3$ mapped onto SCHU S4 is shown in figure 4-7b. The deletion corresponds to loss of a repeated biotin-binding repeat region, leaving *holarctica* strains with two biotin-binding domains while the *tularensis* strains contain three. Whether the deletion occurred during the translocation event and whether it affects function of the pyruvate dehydrogenase complex is not clear. The AceF orthologues from several pathogenic species, including *Vibrio*, *Yersinia*, *Shigella* and *E. coli*, carry three domains. It has, however, been shown that deletion of two of the three domains of AceF in *E. coli* has little affect on function [106]. Whether the additional binding domain influences efficiency of the reaction and whether it could contribute to virulence will require further experimentation. It is worth noting that the *aceF*

truncation was also identified by comparative genome hybridization studies [14, 90, 91], but the translocation event corresponding to CR3 was not detected, underscoring the importance of using multiple approaches for comparative genome analyses.

In addition to direct disruption as a consequence of translocation, genes near the junctions of the translocation events could also be subject to control by unique regulatory machinery. In this light, it is interesting to note that some of the genes within the CR and near the junctions could have functions related to physiology and virulence of *F. tularensis*. Two different genes encoding pilin subunits (*pilE* homologues) of a type IV pilus, are present within the CR2 and CR10, and fine-structural genome comparisons of these $CR_{holarctica}$ mapped onto SCHU S4 are shown in figures 4-7a and 4-7c, respectively. The gene encoding the pilin subunit *pilE5* (FTT0230c) [107] is embedded within CR2. Another member of the *pilE* family is also present in CR10 and the region upstream is disrupted by IS elements. These IS elements are also associated with disruption and duplication of the FTT1311gene in LVS as compared to a single, intact copy in SCHU S4.

Metabolic genes besides *aceF* were also found within the CR; and in particular, several genes encoding enzymes associated with glycerol fermentation are present within the CR11, CR13, and CR15. Fine-structural genome comparisons of each of these $CR_{holarctica}$ mapped onto SCHU S4 are shown in figures 4-7d, 4-7e, and 4-7f, respectively. The inability to ferment glycerol is a hallmark of the *holarctica* subspecies, so it will be interesting to determine if the unique organization involving these three CRs within each subspecies contributes to their different glycerol fermentation phenotypes.

4.3 PESM Discussion:

Whole genome sequencing has provided an outstanding resource for comparative genome studies, allowing high-resolution snapshots of the genetic diversity found within a given species. One of the drawbacks of comparative genome sequencing, however, is that only limited numbers

of strains or taxa can reasonably be compared, making it difficult to distinguish between strain-level genomic differences and true lineage or population-specific sets of genes. Comparative genome hybridization using DNA microarrays circumvents this problem to some degree by providing a single platform for comparison of multiple strains or taxa. On the other hand, the array approach is limited to assessing the diversity in genetic content that is represented on the array. Here, we have shown that PESM can help circumvent the strain bias and the representation problems associated with whole genome sequencing and CGH microarrays. PESM was originally developed as a means to identify unique genomic islands [105]. In our application, we scaled PESM for comparative genome studies by combining the resolution of comparative paired-end sequencing with the power of multiple strain comparison. Given at least one reference genome, PESM provides an economical means to identify candidate regions of genomic difference, and these regions can be further examined in larger strain sets by nested CG-PCR. The PESM library used in our study carried modest sized fragments of the genome (averaging 14 Kb), such that coverage could be obtained with a reasonable amount of sequencing without severely limiting the ability to measure physical size. PESM libraries, however, can be made using different insert sizes in different types of vectors, such that coverage per clone can be increased with larger segments while resolution can be increased by sequencing a larger number of segments from small-fragment libraries.

In addition to economy, the PESM approach allows any strain to be used as a source of the library, thereby allowing the user to choose the best taxonomic unit as a reference. This is particularly important when multiple subpopulations of a species may display unique characteristics that are of interest. Indeed, although we have used PESM in a binary comparison (*holarctica* compared to *tularensis*), it is possible to scaffold multiple libraries into the same PESM pipeline using only a single reference genome upon which to scaffold the data.

4.3.1 Genome Diversity In *F. tularensis*.

Populations of highly virulent bacteria display a wide spectrum with respect to genetic diversity. On the one hand, populations of species such as *Bacillus anthracis* display very little diversity and only a limited number of clones appear to be spread worldwide [74]. These clones can only be differentiated by examining variation in tandem repeats, which are some of the most rapidly evolving loci in the genome [108]. With the availability of multiple genome sequences, single nucleotide polymorphisms will soon complement or displace the MLVA-based approaches. At the other end of the spectrum are subpopulations of *E. coli* O157:H7 which, despite the presence of highly clonal signatures in their genomic backbone, display substantial genomic diversity, even being detectable by a relatively low-resolution method as Pulsed Field Gel Electrophoresis [72, 109, 110].

Based on the data described in our study, we believe that *Francisella tularensis* may represent an intriguing model of genome evolution. Previous studies of genetic diversity in *F. tularensis* detected only limited diversity [13, 15]. The four *F. tularensis* subspecies are known to share 98% identity in their 16S rRNA, show very similar biochemical profiles, and have quite similar antigenic compositions [13, 79]. Only very high resolution methods can provide any phylogenetic signal that reasonably correlates with biochemical and virulence characteristics.

Despite the apparently limited degree of genetic diversity, the *F. tularensis* subspecies display quite distinct geographic distribution and virulence characteristics. Thus, it was initially believed that although limited, the diversity in genomic content would parallel phylogeographic and epidemiologic characteristics and provide clues to the genetic basis for these traits. With the exception of differences in numbers of *pilE*-like loci [107] and the loss of the *pdpD* locus [55], no other obvious candidate virulence genes [10, 48] emerged from comparative genome hybridization studies [14, 90, 91], and >99% of the genetic material present in the more virulent subspecies *tularensis* can also be found in the *holarctica* genome.

In the present study, we now show that despite the high degree of genetic conservation, the genome organization of the *tularensis* and *holarctica* subspecies is vastly different. At least 17 substantial genomic events have occurred during divergence of these two subspecies and have been preserved among multiple strains of each population. The events correspond to extensive translocations and rearrangements, many if not all of which were mediated by movement of IS elements. As shown in Fig. 3, the IS elements are plentiful in the SCHU S4 genome, with 50 different copies of ISftu1 and 16 copies of ISftu2 being distributed around the genome [48]. Given the large number of these elements, it is therefore not surprising to find them at or near the junctions of all 17 CR. Certainly, IS elements were also found abutting subspecies-specific regions of genomic difference (RD) observed in comparative genome hybridization studies [90, 91], and our data here further confirm that IS elements are the primary means through which this genome diversifies.

While the degree of diversity in organization between the genomes of subsp. *tularensis* and subsp. *holarctica* is remarkable, perhaps equally remarkable is the degree to which the unique structure is preserved across temporally and spatially distinct taxa of *holarctica* strains. This observation leads to several interesting possible hypotheses. First, it is possible that population growth is very minimal such that little diversity has had time to accrue. However, because *Francisella* is free-living and is also capable of infecting many different mammalian hosts, slow population turnover in the environment would seem to be an unlikely explanation. A second explanation is that IS elements move only at a very low frequency, thus generating diversity only on a very slow timescale. In this instance, the divergence would have been quite ancestral given the degree of diversity that has accrued. With the number of ISftu1 and ISftu2 elements in the genome, this explanation is unsatisfying. Moreover, we detected significantly more diversity among the $CR_{holarctica}$ regions within the subsp. *tularensis* strains, suggesting that the ISftu are indeed functional. Lastly, and more likely, it is also possible that the extant populations of *holarctica* are quite homogenous because they share very recent common ancestry. This

hypothesis would imply that the populations have recently been through periodic selection or they arose from recent emergence, expansion, and geographic spread of a successful clone, which may also explain why the recently emerged *holarctica* population can be found in Eurasia whereas the *tularensis* populations seem confined to N. America.

4.3.2 The *holarctica* Subspecies Is Likely A Derived State.

In the midst of limited genetic diversity, the simplest explanation for the observed population structure of the *F. tularensis* subspecies is that they are essentially clonal populations and share a common ancestor. Given the high degree of virulence that is displayed by the *tularensis* subspecies, it has been speculated that it represents the ancestral state while the less virulent subspecies are derived states [14] that are more adept at infecting hosts without killing. In support of this hypothesis, the *novicida* subsp. can be found in water, implying that it may survive more effectively in the free living state than the more highly virulent *tularensis* subspecies. Evolutionary analysis of VNTR loci also suggest that *tularensis* is likely more similar to the common ancestor [2, 76, 78]. With respect to genome organization, our data also support this hypothesis, showing that organization of the different $CR_{holarctica}$ within subsp. *holarctica* appears to be a derived state, arising by dissociation of genomic units through translocation events in an immediate ancestor of the *holarctica* populations. At least three genomic segments were found to be single contigs in the tularensis genome but are dispersed into six different CR in the *holarctica* subsp. Moreover, some genes at the junctions of these events are disrupted or even deleted in *holarctica* whereas the respective genes are present with no remnants of gene fragments being present at the junctions of *tularensis*. Furthermore, the disruption of the apparent glycerol fermentation operon through translocation in *holarctica* is likely the derived condition. We also note that three additional CR in tularensis (CR3-CR9, CR4-CR8, and CR9-CR11) are adjacent, but not contiguous whereas they are highly dispersed in the *holarctica* subspecies. Therefore, evidence is beginning to mount in favor of the hypothesis that

F. tularensis subsp. *tularensis* is likely more similar to the ancestral state while populations of the *holarctica* subsp. are derived states. If this is true, then analysis of genomic content and organization between the different subspecies should provide insights not only into additional candidate virulence loci, but also into selective pressures that have led to emergence and geographical spread of the *holarctica* populations.

4.3.3 Two US Strains May Represent A Unique Taxonomic Unit Within Francisella.

Although our search was primarily focused on identifying population-specific regions of genomic difference, the genome organization observed in the subsp. *tularensis* strains WY00W4114 and WY-WSLVL02 is very intriguing. Their pattern of genome organization is clearly distinct from the *tularensis* and *holarctica* populations, sharing CR "alleles" at some loci with *tularensis* strains, CR "alleles" at other loci with *holarctica* strains, and unique alleles at still other loci. The diversity is such that we propose they represent a new taxonomic unit, and this will be further discussed in chapter 5.

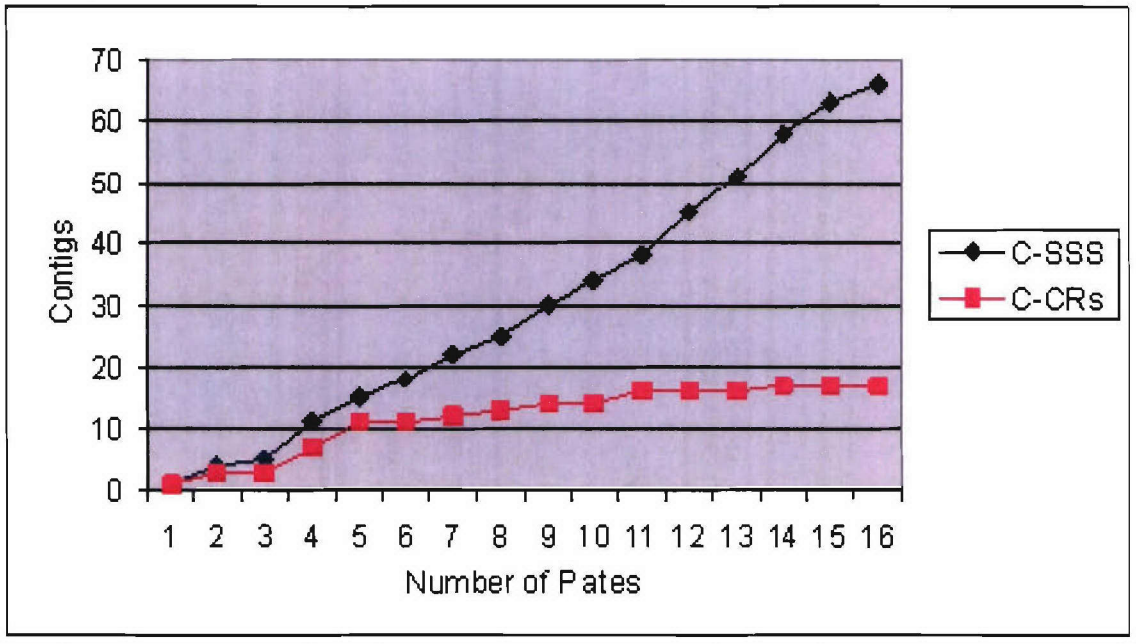
Figure 4-1.

Statistical Results of Plaquing Experiments, Generation of Successful Direct Plaque Amplicons, and Cumulative $CR_{holarctica}$ Segments.

Shown across the table from left to right for each plaquing experiment (designated by plate number) are the cumulative (C) subspecies-specific segments (SSS), cumulative CR count, successful (S) direct-plaque amplicon (DPA) count, and cumulative DPA successes. The total counts for each column are shown, as are the average DPA clone size (~14.17 kb), total number of DPA clones obtained (n=752) vs. the number of successful DPA obtained (n=551), the overall efficiency of S-DPA to total DPA (~73%), and approximate genome coverage (~4.1 X). The chart (bottom panel) shows the total number of subspecies-specific segments (contigs) plotted against the cumulative number of CRs obtained after all plates were processed. Note that the last new CR was obtained from plate-15 and plateaued after that point.

Figure 4-1: Statistical Results of Plaquing Experiments.

Plate #:	C-SSS	C-CRs	S-DPA/Plate	C-DPA Successes
1	1	1	28	28
2	4	3	30	58
3	5	3	38	96
4	11	7	44	140
5	15	11	42	182
6	18	11	31	213
7	22	12	26	239
8	25	13	19	258
10	30	14	41	299
11	34	14	39	338
12	38	16	43	381
13	45	16	25	406
14	51	16	38	444
15	58	17	34	478
16	63	17	38	516
17	66	17	35	551
T = 16	T = 66	T = 17	T = 551 S-DPA	T=551 DPA Successes
Average DPA Clone = 14,174 bp			Total of 752 DPA Clones obtained	
Efficiency= ~73.3%; Genome Coverage = ~4.11 X based on 1.9 Mb Genome Size				



Legend: C=Cumulative, SSS=Subsp.-spec. Segment, S=Successful, DPA=Direct Plaque Amplicon

Figure 4-2.

Direct-Plaque Amplification (DPA) PCR Experiment.

The figure shows the respective bands for clones #1-47 of plaquing experiment #10 following DPA-PCR. The 1-15 kb size-standard molecular ruler (each band=1 kb) lanes are as marked. The size-standard lanes allowed for accurate size determinations of each DPA-clone amplicon.

Figure 4-2: Direct-Plaque Amplification (DPA) PCR Experiment.

Figure 4-3.

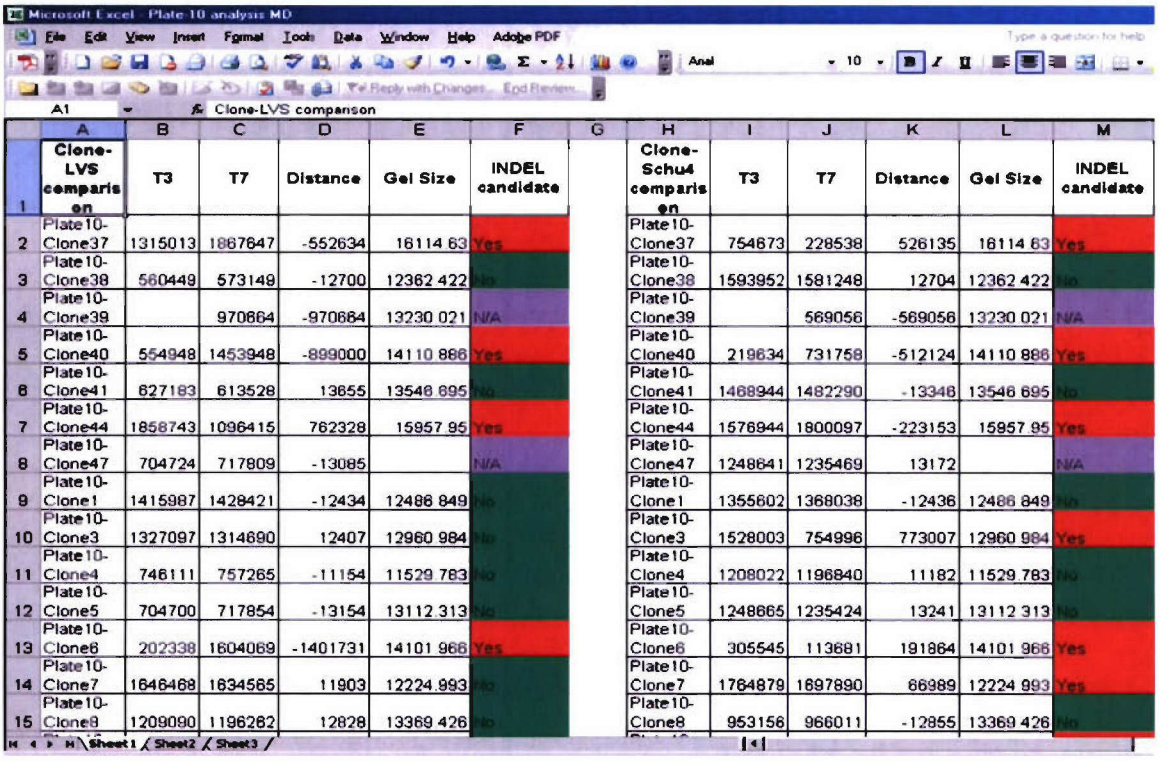
Raw and Sorted PESMP Data for Identification of $CR_{holarctica}$ Segments.

Panel 4-3a shows raw PESMP data output into an Excel spreadsheet following input of sizing and sequencing data into the PESM Pipeline of the UNL Pathogene Server. The spreadsheet shows data of sequence from each of the T3 and T7 primers and size data (determined from gel electrophoresis) from each clone compared with its corresponding sequence coordinates and resultant size (between coordinates) from the LVS WGS (left-hand side) and SCHU S4 WGS (right-hand side). Quality clones (having sequence from both ends and a visible amplicon band) having size agreement within ~+/- 2 kb and are shown in green highlighting. Quality clones with non-congruent sizes are shown in red highlighting. Clones not meeting "Quality" standards are shown in lavender highlighting. Clones having size congruence with both LVS and SCHU S4 were considered to be *F. tularensis* species-specific whereas those in agreement with LVS but not SCHU S4 were considered subspecies-specific candidates and grouped into the $CR_{holarctica}$ segments.

Panel 4-3b shows sorted PESMP data from the raw Excel spreadsheet for identification of the $CR_{holarctica}$ contigs. The screen shown is just one screen of the composite table from all 16 sequencing and sizing plates. The entire table was sorted according to the correlation of T3 – T7 coordinates with the LVS WGS coordinates. This sorting strategy resulted in identification of the seventeen $CR_{holarctica}$ segments positioned clockwise, beginning at 12 o'clock around the LVS WGS (shown in figure 4-6).

Figure 4-3: Raw and Sorted PESMP Data for Identification of $CR_{holarctica}$ Segments.

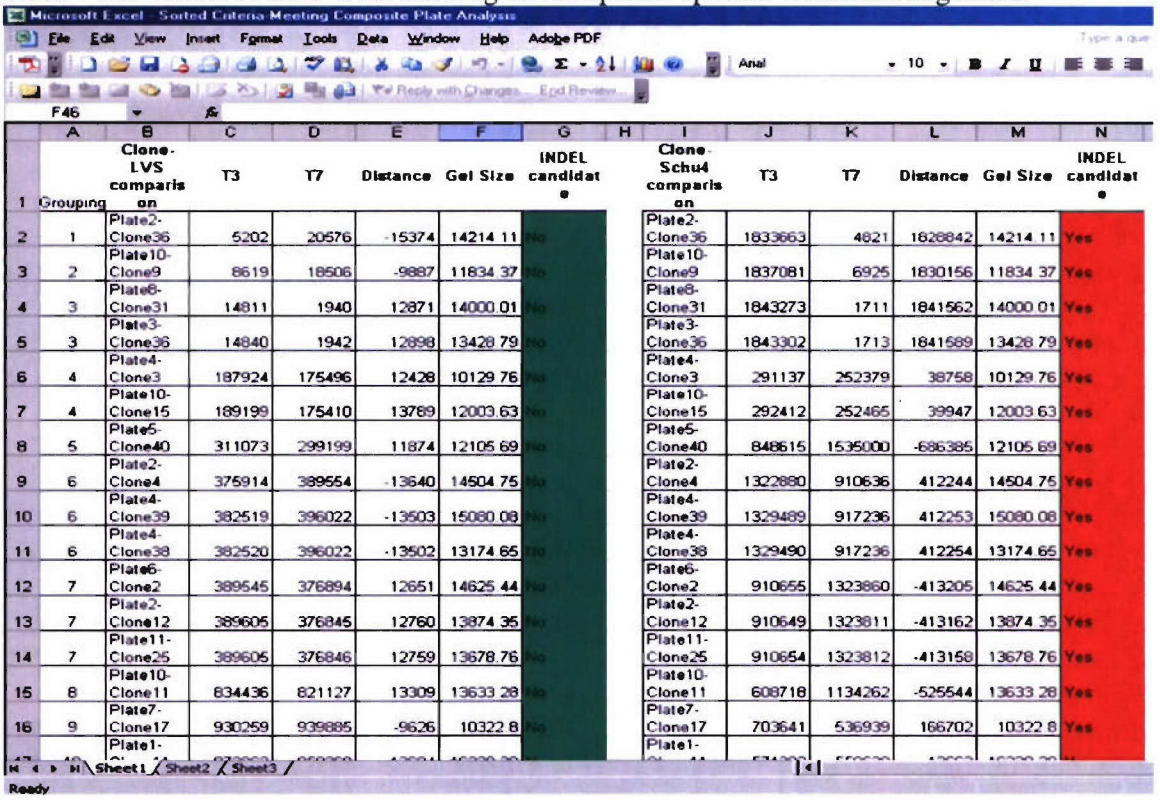
Panel 4-3a: Raw PESMP Data.



Microsoft Excel - Plate 10 analysis MD

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Clone-LVS comparison	T3	T7	Distance	Gel Size	INDEL candidate		Clone-Schu4 comparison	T3	T7	Distance	Gel Size	INDEL candidate
2	Plate10-Clone37	1315013	1867647	-552634	16114.63	Yes		Plate10-Clone37	754873	228538	526135	16114.63	Yes
3	Plate10-Clone38	560449	573149	-12700	12362.422	No		Plate10-Clone38	1593952	1581248	12704	12362.422	No
4	Plate10-Clone39		970664	-970664	13230.021	N/A		Plate10-Clone39		569056	-569056	13230.021	N/A
5	Plate10-Clone40	554948	1453948	-899000	14110.886	Yes		Plate10-Clone40	219634	731758	-512124	14110.886	Yes
6	Plate10-Clone41	627183	613528	13655	13546.695	No		Plate10-Clone41	1468944	1482290	-13346	13546.695	No
7	Plate10-Clone44	1858743	1096415	762328	15957.95	Yes		Plate10-Clone44	1576944	1800097	-223153	15957.95	Yes
8	Plate10-Clone47	704724	717809	-13085		N/A		Plate10-Clone47	1248641	1235469	13172		N/A
9	Plate10-Clone1	1415987	1428421	-12434	12486.849	No		Plate10-Clone1	1355602	1368038	-12436	12486.849	No
10	Plate10-Clone3	1327097	1314690	12407	12960.984	No		Plate10-Clone3	1528003	754996	773007	12960.984	Yes
11	Plate10-Clone4	746111	757265	-11154	11529.783	No		Plate10-Clone4	1208022	1196840	11182	11529.783	No
12	Plate10-Clone5	704700	717854	-13154	13112.313	No		Plate10-Clone5	1248865	1235424	13241	13112.313	No
13	Plate10-Clone6	202338	1604069	-1401731	14101.966	Yes		Plate10-Clone6	305545	113681	191864	14101.966	Yes
14	Plate10-Clone7	1646468	1634565	11903	12224.993	No		Plate10-Clone7	1764879	1697890	66989	12224.993	Yes
15	Plate10-Clone8	1209090	1196262	12828	13369.426	No		Plate10-Clone8	953156	966011	-12855	13369.426	No

Panel 4-3b: Sorted PESMP Data showing all Subspecies-specific Candidate Segments.



Microsoft Excel - Sorted Criteria-Meeting Composite Plate Analysis

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Grouping	Clone-LVS comparison	T3	T7	Distance	Gel Size	INDEL candidate		Clone-Schu4 comparison	T3	T7	Distance	Gel Size	INDEL candidate
2	1	Plate2-Clone36	5202	20576	-15374	14214.11	No		Plate2-Clone36	1833663	4821	1828842	14214.11	Yes
3	2	Plate10-Clone9	8619	18506	-9887	11834.37	No		Plate10-Clone9	1837081	6925	1830156	11834.37	Yes
4	3	Plate8-Clone31	14811	1940	12871	14000.01	No		Plate8-Clone31	1843273	1711	1841562	14000.01	Yes
5	3	Plate3-Clone36	14840	1942	12898	13428.79	No		Plate3-Clone36	1843302	1713	1841589	13428.79	Yes
6	4	Plate4-Clone3	187924	175496	12428	10129.76	No		Plate4-Clone3	291137	252379	38758	10129.76	Yes
7	4	Plate10-Clone15	189199	175410	13789	12003.63	No		Plate10-Clone15	292412	252465	39947	12003.63	Yes
8	5	Plate5-Clone40	311073	299199	11874	12105.69	No		Plate5-Clone40	848615	1535000	-686385	12105.69	Yes
9	6	Plate2-Clone4	376914	389554	-13640	14504.75	No		Plate2-Clone4	1322880	910636	412244	14504.75	Yes
10	6	Plate4-Clone39	382519	396022	-13503	15080.08	No		Plate4-Clone39	1329489	917236	412253	15080.08	Yes
11	6	Plate4-Clone38	382520	396022	-13502	13174.65	No		Plate4-Clone38	1329490	917236	412254	13174.65	Yes
12	7	Plate6-Clone2	389545	376894	12651	14625.44	No		Plate6-Clone2	910655	1323860	-413205	14625.44	Yes
13	7	Plate2-Clone12	389605	376845	12760	13874.35	No		Plate2-Clone12	910649	1323811	-413162	13874.35	Yes
14	7	Plate11-Clone25	389605	376846	12759	13678.76	No		Plate11-Clone25	910654	1323812	-413158	13678.76	Yes
15	8	Plate10-Clone11	834436	821127	13309	13633.28	No		Plate10-Clone11	608718	1134262	-525544	13633.28	Yes
16	9	Plate7-Clone17	930259	939885	-9626	10322.8	No		Plate7-Clone17	703641	536939	166702	10322.8	Yes
		Plate1-							Plate1-					

Table 4-1.

$CR_{holarctica}$ Coordinates in LVS and Corresponding SCHU S4 Coordinates.

The table shows the corresponding LVS WGS coordinates as well as the SCHU S4 subsegment sequence coordinates for each $CR_{holarctica}$. Navy blue SCHU S4 coordinates indicate same synteny with corresponding LVS sequence; red coordinates indicate an inversion of the sequence segment between the two genomes; and green coordinates indicate segments with sequence homology to IS elements (primarily ISftu1 and ISftu2). The IS-element homolog coordinates are beneath the main sequence coordinates at positions indicating where they are found between SCHU S4 sequence subsegments.

Table 4-1: $CR_{holarctica}$ Coordinates in LVS and Corresponding SCHU S4 Coordinates.

CR #:	LVS Coordinates	SCHU S4 Subsegment Coordinates (Red=Inverted) (Green=Inserted ISE homologs)
1	1940 - 20576	1712-4041 1831793-1845643 8297-4841
2	174549 - 189301	253326-242638 287498-292513
3	299160 - 311115	1535039-1531661 839152-848615
4	375912 - 396023	1322878-1331182 324776-323900 // 907086-917236 1476864-1475990
5	431097 - 443907	391758-403448 // 12596-12811 606879-607749
6	821127 - 834436	1134262-1132436 501838-501715 1132449-1122289 607742-608718
7	930259 - 939885	703641-711588 534357-536939
8	1225017 - 1235761	937237-932599 357261-362164
9	1314561 - 1327233	755134-753179 1516527-1528130
10	1379753 - 1403782	792818-784939 // 1332861-1335918 // 1332861-1335918 // 1335916-1343388 1577105-1576237 607688-606878 1371321-1370447
11	1429245 - 1442913	1368862-1371320 738532-749777
12	1471577 - 1482382	713935-710637 535308-526858
13	1563868 - 1580323	1659170-1649695 144417-138349 137410-136423
14	1634452 - 1648610	1697768-1708621 287501-288454 1767007-1764746
15	1687717 - 1699637	148626-143466 103008-953005
16	1780999 - 1797792	14808-12996 // 12998-8922 1801624-1811724 1370447-1371321
17	1850023 - 1863814	194767-192501 1911892-185921 219083-224690

Figure 4-4.

PESM *Francisella* Panel CG-PCR Run.

Shown in the figure is a representative CG-PCR run, and in this case, for the CR-10 locus. The DNA samples are listed from #s 1-92 according to the AFIP-UNMC *Francisella* panel (Table 2-1). Positions #94 and #95 (numbers not shown on gel) are for the SCHU S4 and LVS DNA controls, respectively; whereas #94 (number not shown) is a repeat of #59. Subsp. *tularensis* CR10-A1-, CR10-A2-, and CR10-A3- (also tentatively called subsp. *neotularensis*) as well as subsp. *holarctica* CR10-B- and subsp. *novicida* CR10-C-sized bands are shown represented by red lettering. The 100 bp size standards (in 100 bp increments up to 3 kb, and with the brightest band corresponding to the 1 kb band) are shown at the outer-most lanes as well as between clones 24 and 25, and between clones 72 and 73.

Figure 4-4: PESM *Francisella* Panel CG-PCR Run.

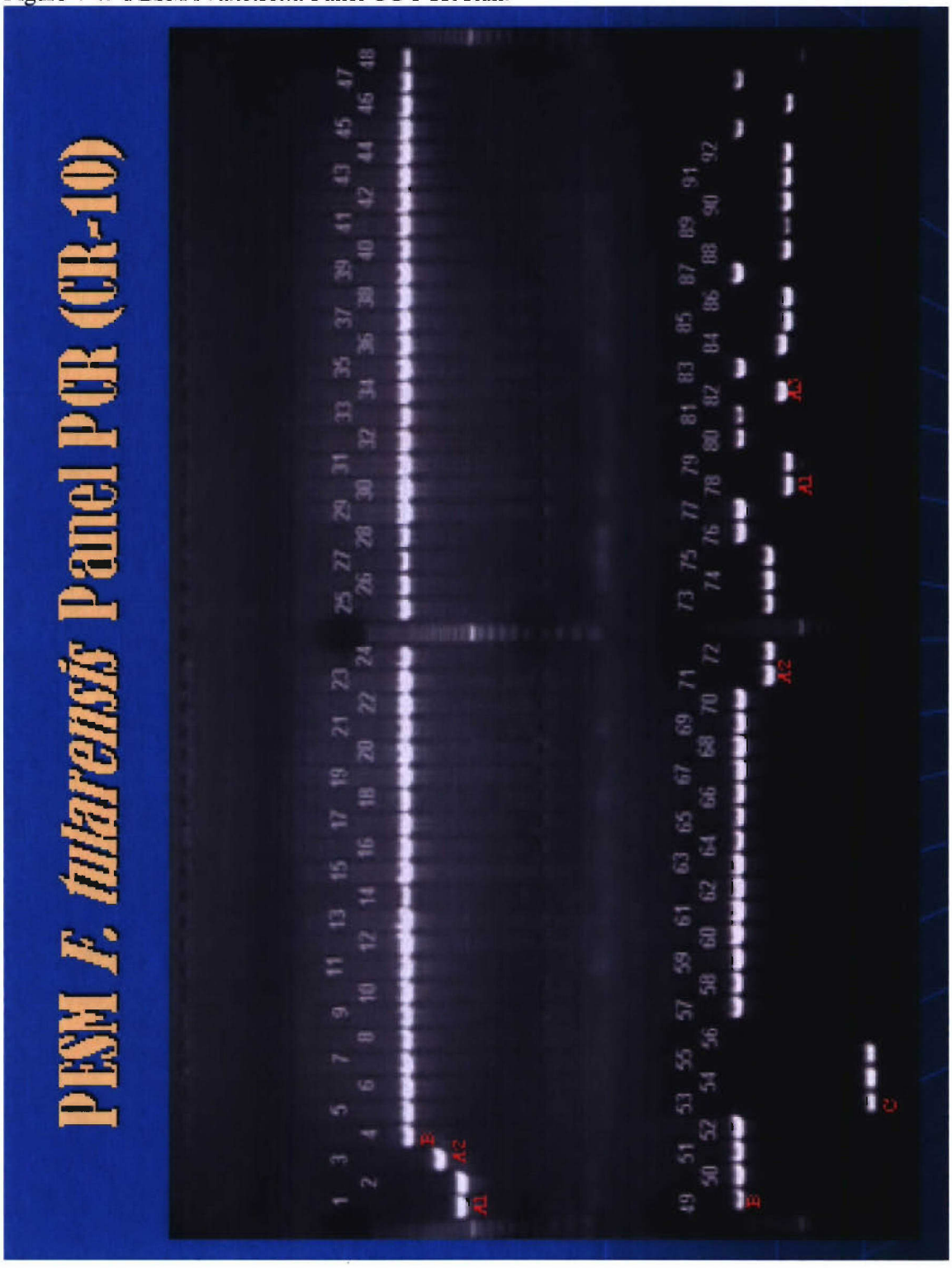


Table 4-2.

Summary of CG-PCR Different Genotypes from AFIP-UNMC *Francisella* Panel.

The table shows groupings of representative genotypes following CG-PCR at all 17 $CR_{holarctica}$ loci for all 93 strains of the AFIP-UNMC Franciesella panel. Note that, although Tu-1 has the same genotype as LVS, it is presented here because it is from the Spanish-outbreak subpopulation (having the RD_{Spain}-deletion) presented in Chapter 3. All negative and/or alternative bands are in green lettering, and the respective band sizes for the latter are listed at the bottom of the table.

Table 4-2: Summary of CG-PCR Different Genotypes from AFIP-UNMC *Francisella* Panel.

Ft Panel Number	Strain-DNA Name or # used in Studies	RD-1 PCR	PESM PCR CR-1	PESM PCR CR-2	PESM PCR CR-3	PESM PCR CR-4	PESM PCR CR-5	PESM PCR CR-6	PESM PCR CR-7	PESM PCR CR-8	PESM PCR CR-9	PESM PCR CR-10	PESM PCR CR-11	PESM PCR CR-12	PESM PCR CR-13	PESM PCR CR-14	PESM PCR CR-15	PESM PCR CR-16	PESM PCR CR-17
	subsp. holarctica																		
4	LVS	Type-B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
5	Tu-1	Type-B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
52	Tu-42	Type-B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	A
	subsp. tularensis																		
2	Schu S4	Type-A	A	A	A	A	A	A	A	A	A	A1	A	A	A	A	A	A	A
86	OK-98041035	Type-A	A	A	A	A	A	A	A	A	A	A1	A	A	A	Neg	A	A	A
93	ATCC 6223	Type-A	A	A	A	A	A	A	A	A	A	A2	A	A	A	A	A	A	A
3	A88R160	Type-A	A	A	A	A	A	A	A	A	A	A2	A	A	A	A	A	A	A
71	88R52	Type-A	A	A	A	A	A	A	A	A	A	A2	A	A	A	A	A	A	A
72	88R144	Type-A	A	A	A	A	A	A	A	A	A	A2	A	A	A	A	A	A	A
73	AK-1133496	Type-A	A	A	A	A	A	A	A	A	A	A2	A	A	A	A	A	A	A
74	AK-1100558	Type-A	A	A	A	A	A	A	A	A	A	A2	A	A	A	A	A	A	A
75	AK-1100559	Type-A	A	A	A	A	A	A	A	A	A	A2	A	A	A	A	A	A	A
93	ATCC-6623	Type-A	A	A	A	A	A	A	A	A	A	A2	A	A	A	A	A	A	A
82	WY-00W4114	Type-A	Neg	B	A	A	A	A	B	A	B	A3	A	B	Neg	B	A	C	A
84	WY-WSVL02	Type-A	Neg	Neg	A	A	A	A	B	A	B	A3	A	B	Neg	B	A	C	A
	F. philomiragia																		
56	*F. philomiragia* 25015	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
	subsp. novicida																		
53	*Ft. novicida* 13482	no viable	Neg	Neg	C	Neg	C	B	C	C	C	C	A	C	C	Neg	B	Neg	C
54	Tu-43	no viable	Neg	C	C	Neg	C	B	C	C	C	C	A	C	C	Neg	B	Neg	C
55	US ARMRID D000506 7002	no viable	Neg	Neg	C	Neg	C	B	C	C	C	C	A	C	C	Neg	B	Neg	C
	subsp. holarctica-japonica																		
59	Japanese	japonica	B	B	B	Neg	B	B	B	Neg	B	B	B	B	D	Neg	B	B	A
	Alternative Bands (bp)			C=2X-1400 1543	C=3X-1581 2400 3600		C-750		C-1200	C-700	C-350	A=3X-A1-1200 A2-1500 A3-1300		C-300	C-1150 D=2X-400 2248			C-1800	C-650

Figure 4-5.

Cumulative PESM CG-PCR Results.

Colored rectangles correspond with CG-PCR results for each of the 17 loci for each of the *Francesella* panel strains tested. Red rectangles correspond with SCHU S4-specific PCR results; green rectangles correspond with LVS-specific PCR results; yellow rectangles correspond with negative PCR results; orange and blue-grey rectangles correspond with PCR reactions unique (not predicted for either SCHU S4 or LVS) for each given locus. Also shown in the figure at the right-hand side is the RD1 PCR result for each strain. Among the Type-A strains (subsp. *tularensis*), the 3 main CR10 genotypes, CR10-A1, CR10-A2, and CR10-A3 (or tentatively, subsp. *neotularensis*) can be observed as denoted by the red, orange, and blue-grey rectangles, respectively. The CR10-A3 genotypic group demonstrates extensive heterogeneity as compared with the other CR10-A genotypes. The Type-B (subsp. *holarctica*) strains clearly demonstrate more homogeneity as shown here. Note that geographic locations as well as other demographic information for the strains may be obtained from Table 2-1 in chapter 2.

Figure 4-5: Cumulative PESM CG-PCR Results.

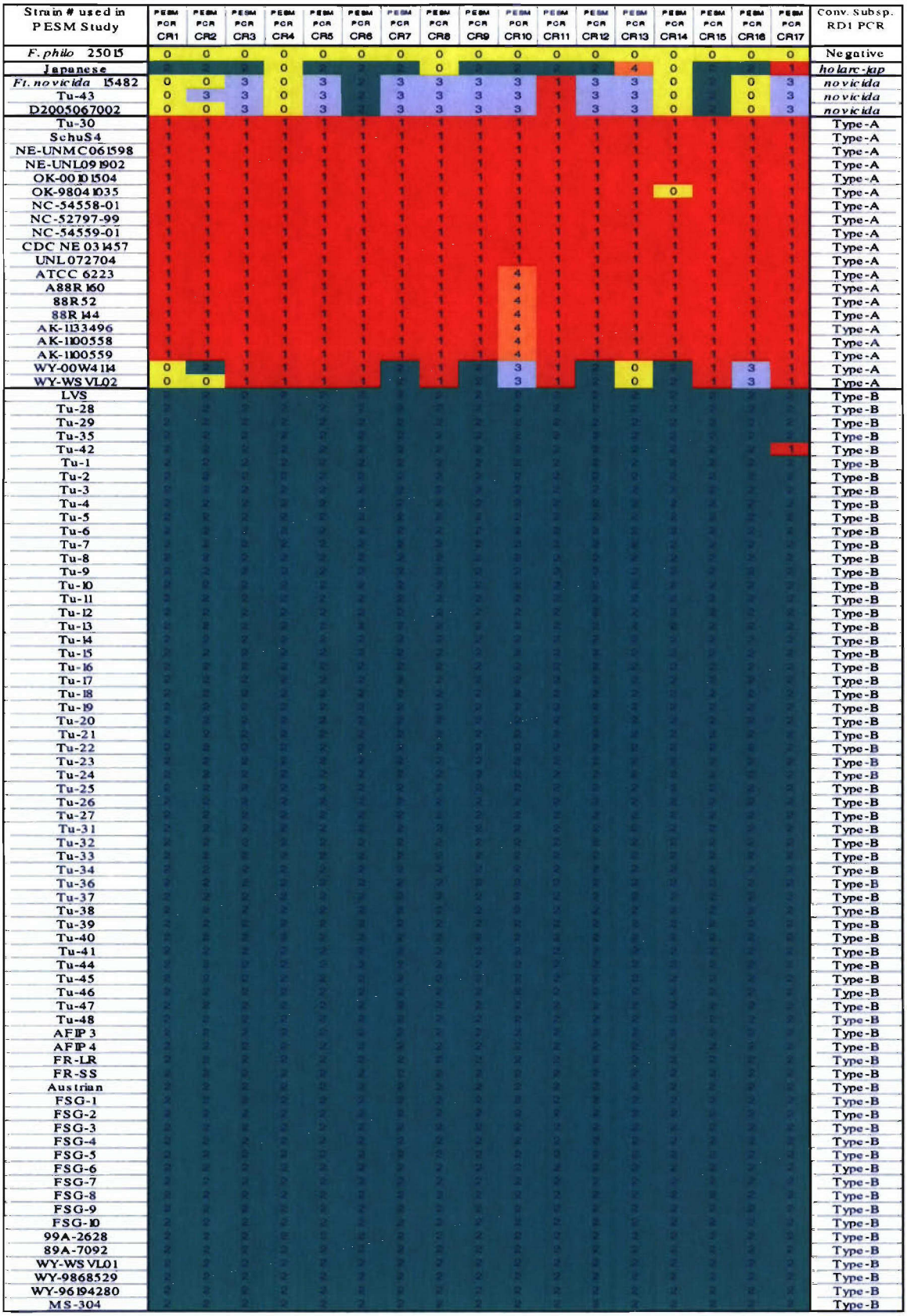


Figure 4-6.

All $CR_{holarctica}$ Mapped Onto The Circular Genome Of *F. tularensis (Ft.)* SCHU S4.

The outer scale designates coordinates in base pairs (bp). The first circle shows predicted coding regions on the plus strand color-coded by role categories: violet, amino acid biosynthesis; light blue, biosynthesis of cofactors, prosthetic groups and carriers; light green, cell envelope; red, cellular processes; brown, central intermediary metabolism; yellow, DNA metabolism; light gray, energy metabolism; magenta, fatty acid and phospholipid metabolism; pink, protein synthesis and fate; orange, purines, pyrimidines, nucleosides and nucleotides; olive, regulatory functions and signal transduction; dark green, transcription; teal, transport and binding proteins; gray, unknown function; salmon, other categories; blue, hypothetical proteins.

The second circle shows the location of all known copies of *F. tularensis* SCHU S4 Isftu1 (grey) and Isftu2 (blue) genes. The fourth circle depicts the genomic location of the 17 *F. tularensis* $CR_{holarctica}$ as indicated by their $CR_{holarctica}$ number. The third circle represents the color-coded matching location of each of the 17 *F. tularensis* $CR_{holarctica}$ distributed onto the genome of *F. tularensis.* SCHU S4.

Figure 4-6: All $CR_{holarctica}$ Mapped Onto The Circular Genome Of *Ft.* SCHU S4.

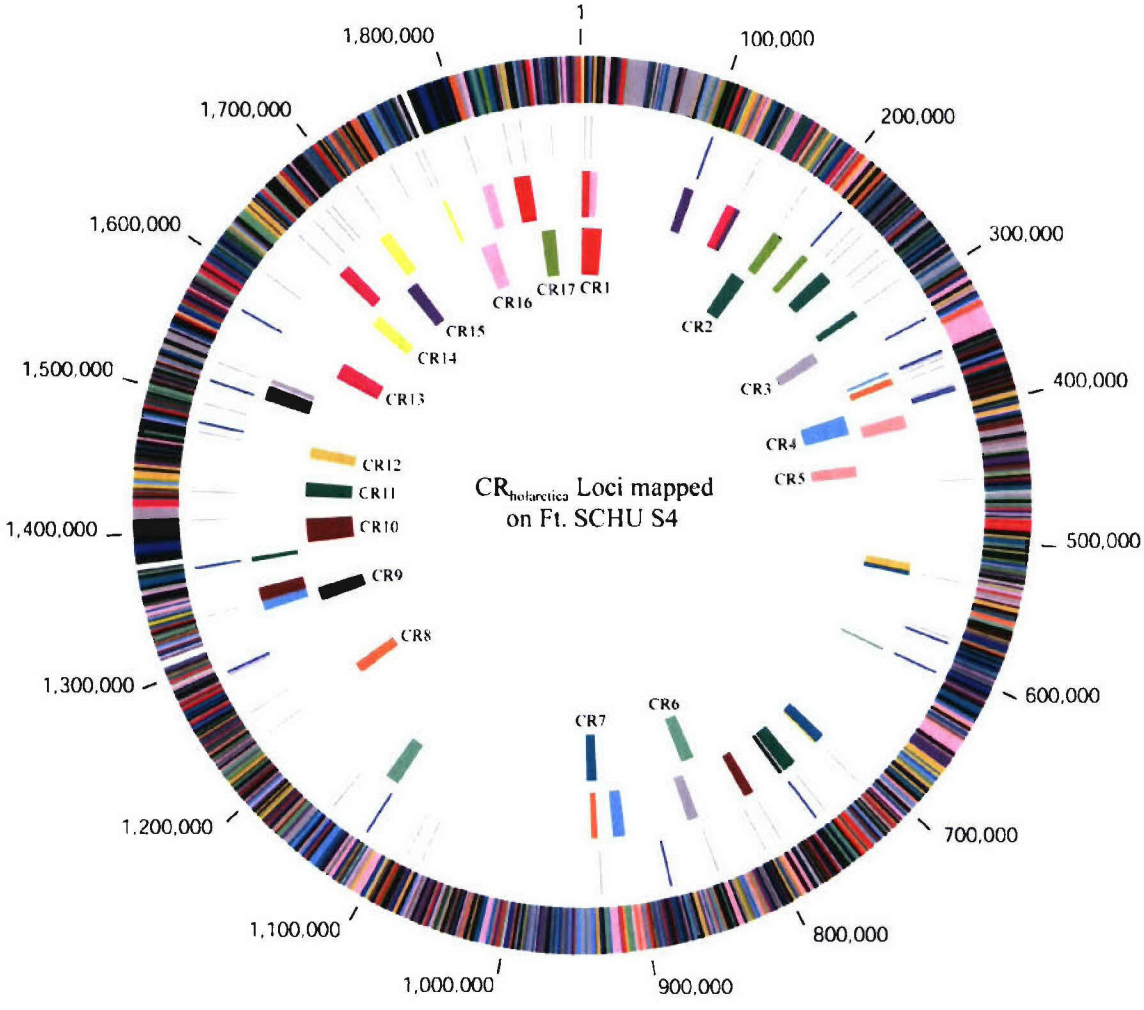
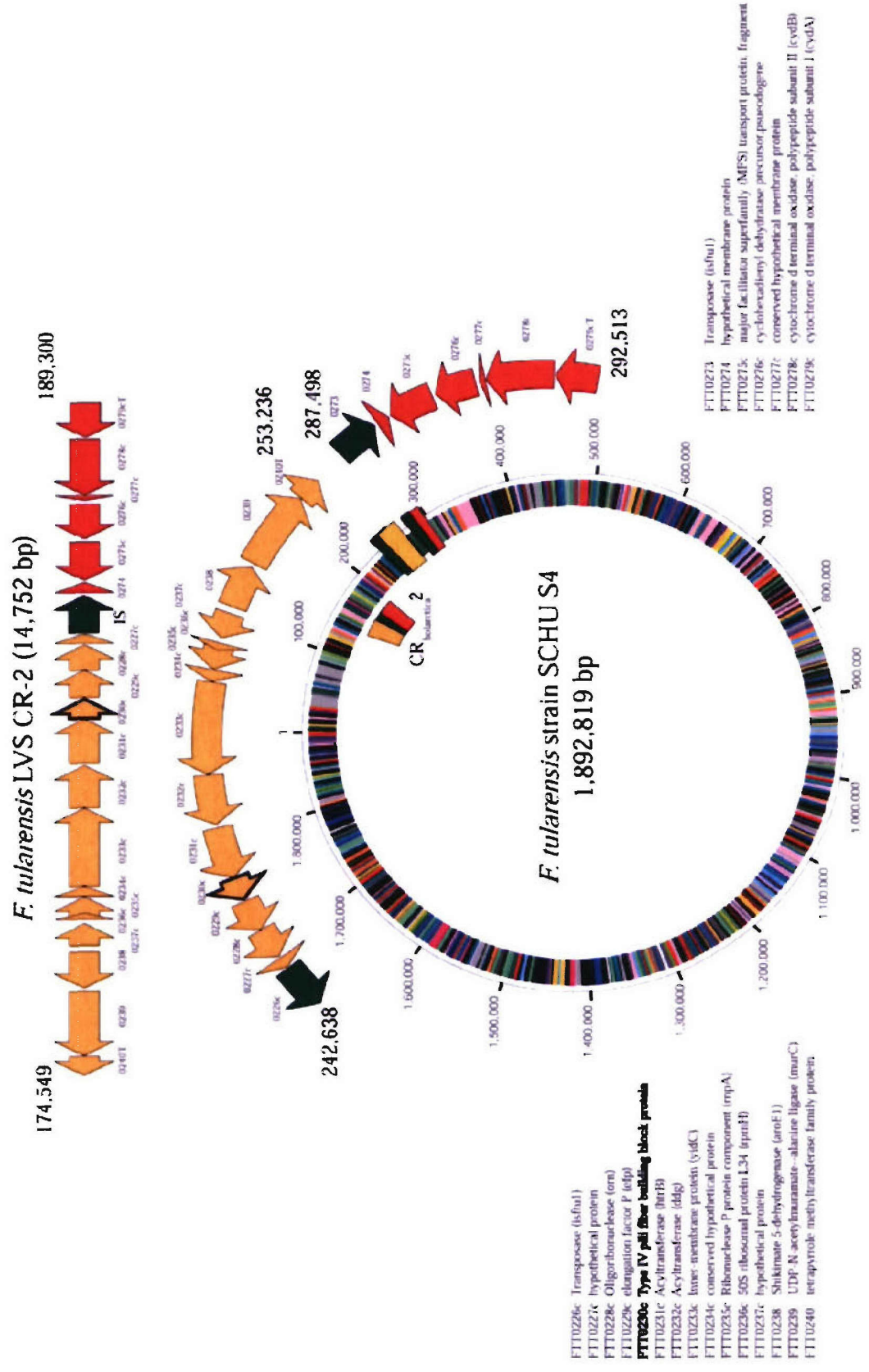


Figure 4-7.

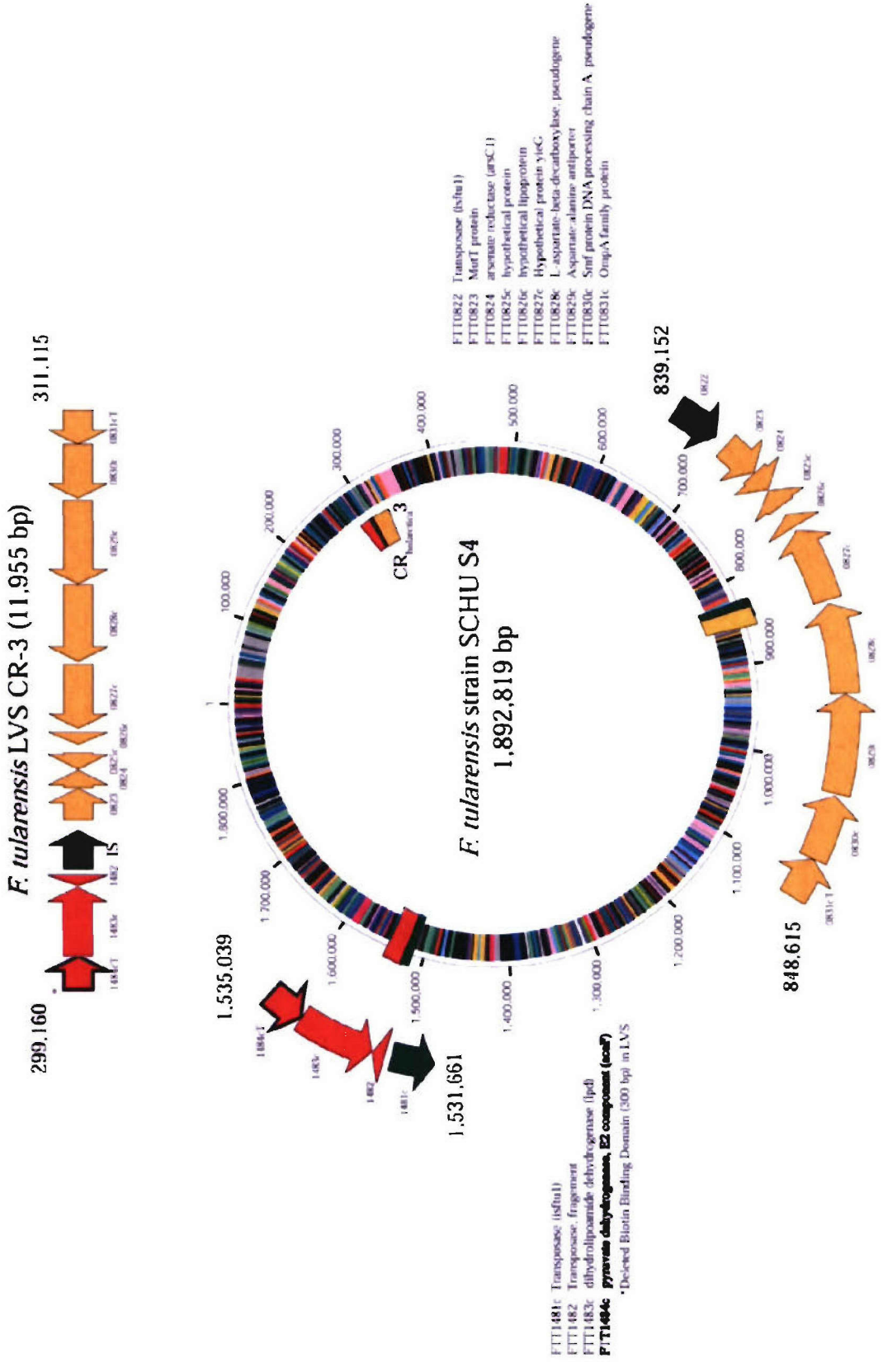
Individual $CR_{holarctica}$ aligned onto Circular *F. tularensis* SCHU S4 Map.

Each $CR_{holarctica}$ for CR2, CR3, CR10, CR11, CR13, and CR15 is shown mapped individually onto the circular SCHU S4 genome in panels 4-7a-f, respectively. Genes or locus tags designated by a prefix or suffix "T" indicate genes truncated due to the beginning or ending of the $CR_{holarctica}$ clone, or due to altered arrangements of genomic structure within SCHU S4 as compared with LVS. Genes (shown as colored arrows) and their corresponding locus tags for virulence- or biochemically-significant genes are bolded, and are presented bolded in the adjacent list of genes for each $CR_{holarctica}$.

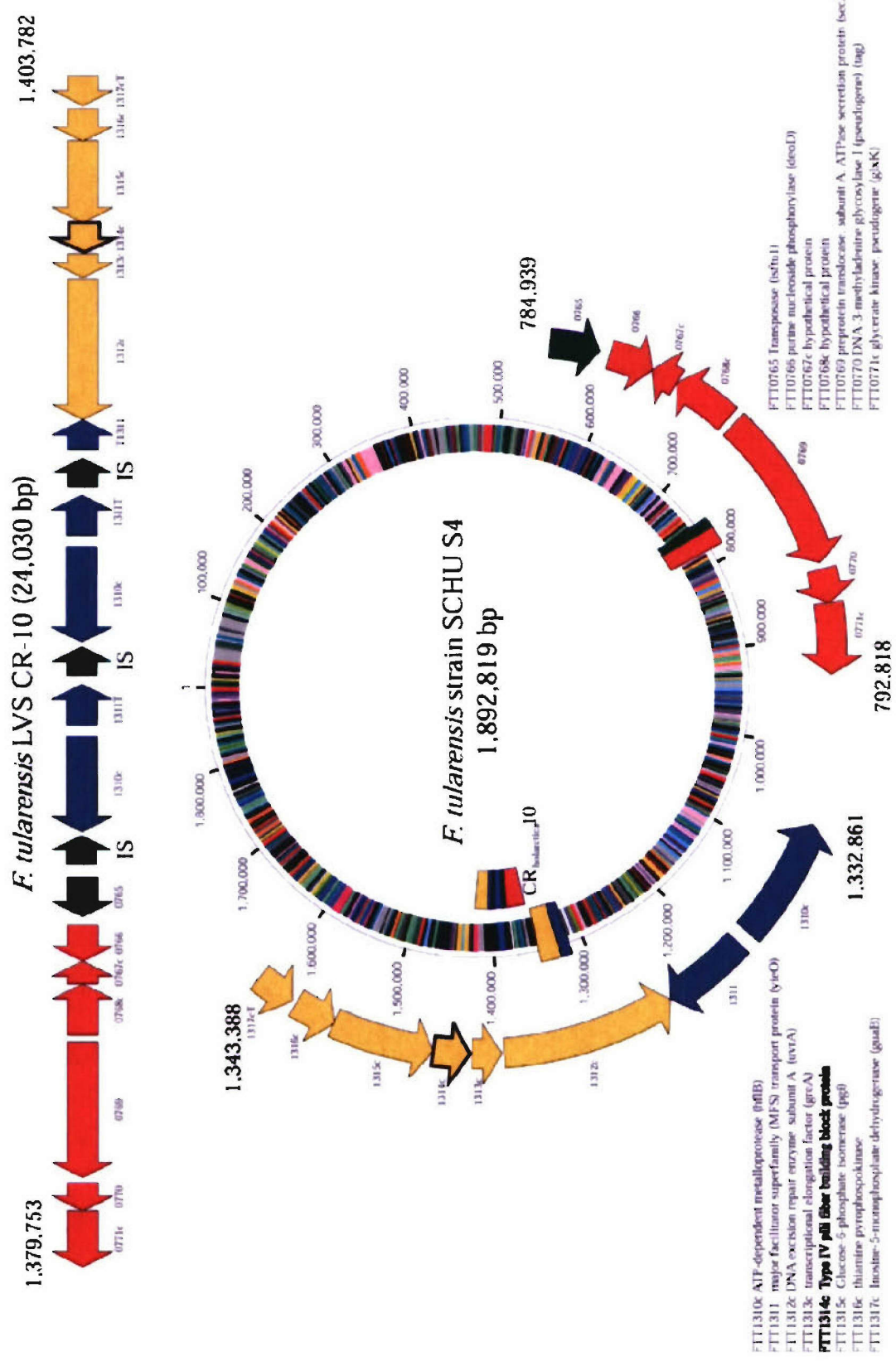
Panel 4-7a: $CR_{holarctica}2$ Aligned Onto The Circular SCHU S4 Map.



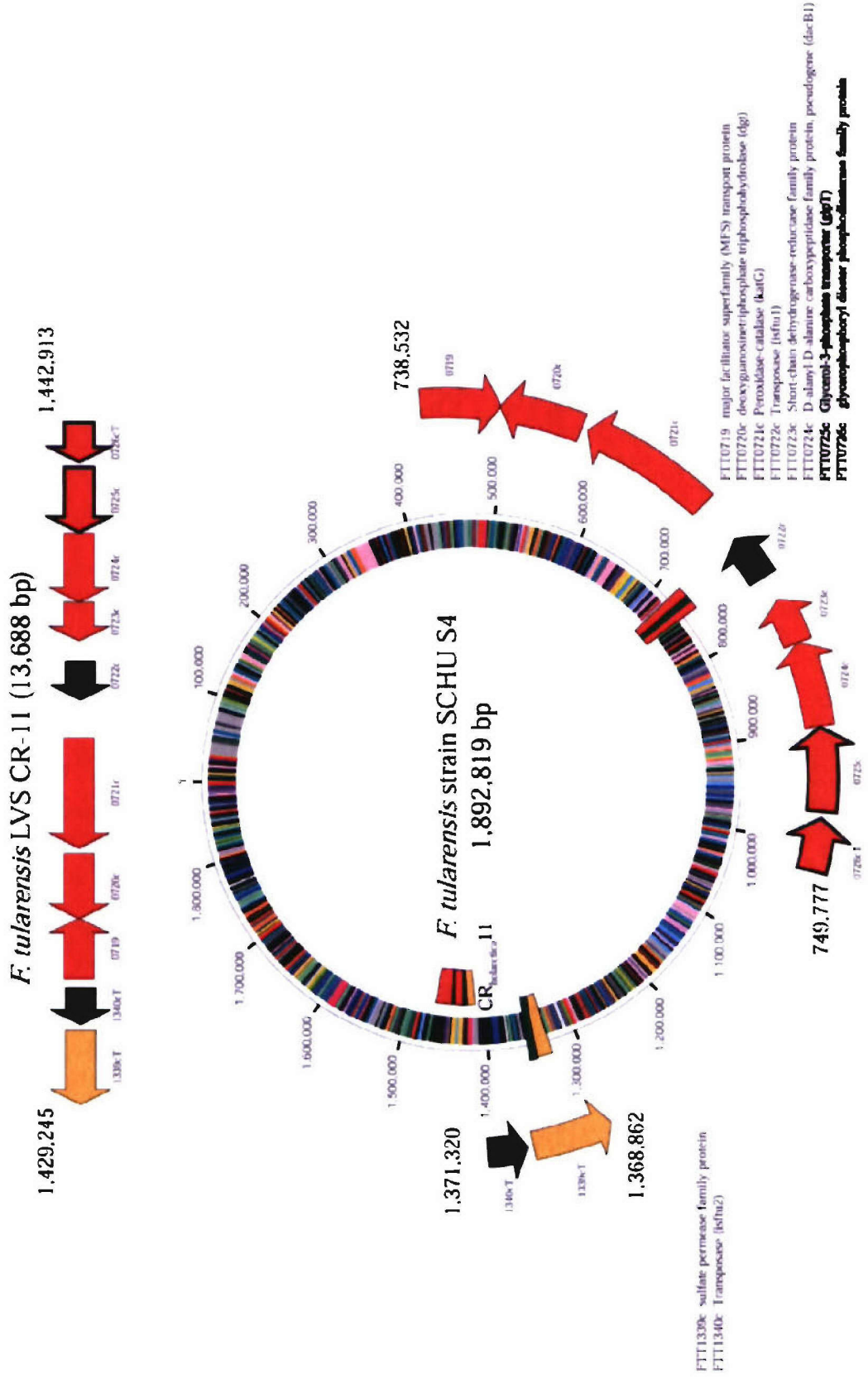
Panel 4-7b: $CR_{holarctica}3$ Aligned Onto The Circular SCHU S4 Map.



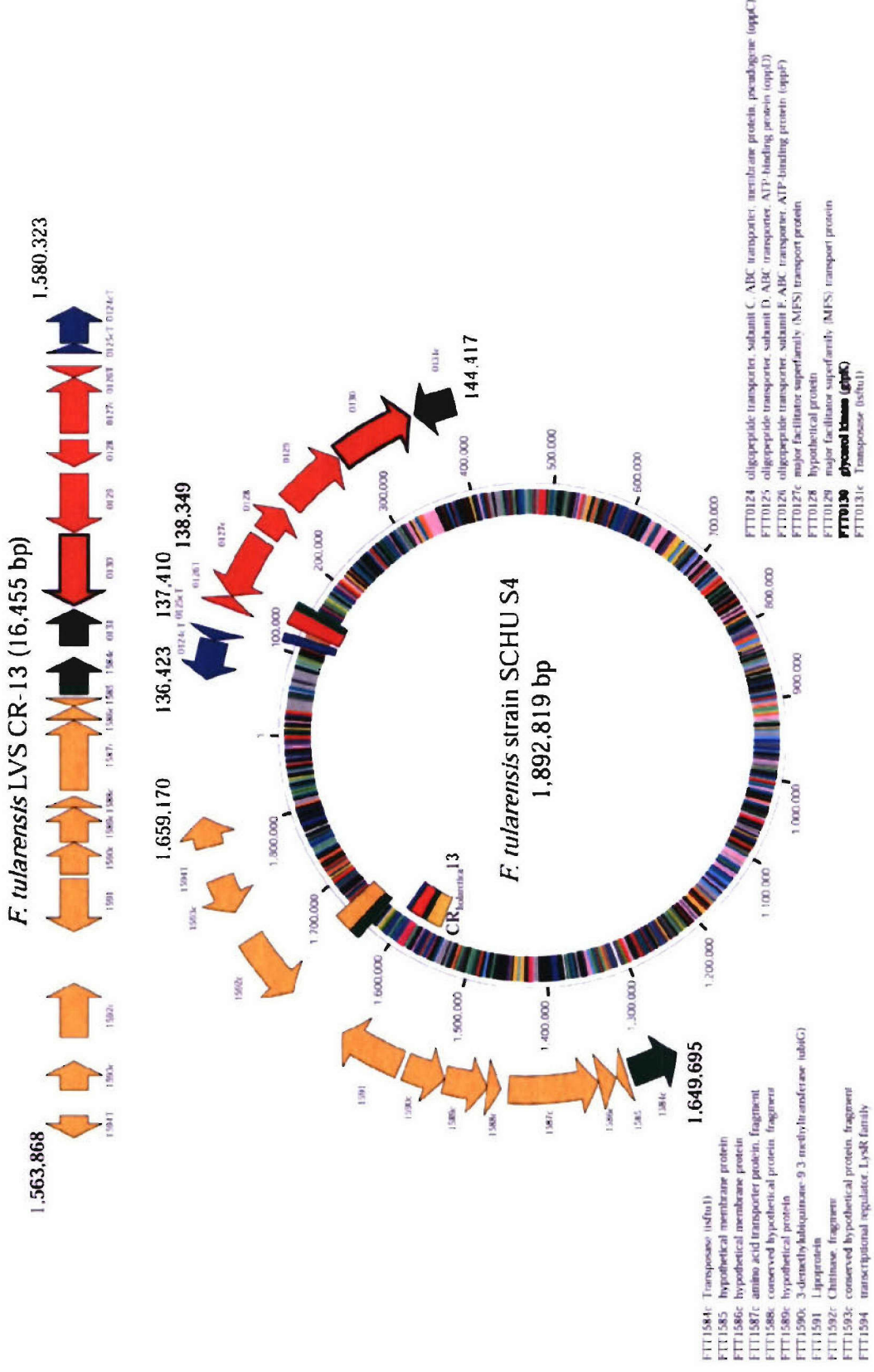
Panel 4-7c: $CR_{holarctica}10$ Aligned Onto The Circular SCHU S4 Map.



Panel 4-7d: $CR_{holarctica}11$ Aligned Onto The Circular SCHU S4 Map.



Panel 4-7e: $CR_{holarctica}13$ Aligned Onto The Circular SCHU S4 Map.



Panel 4-7f: $CR_{holarctica}15$ Aligned Onto The Circular SCHU S4 Map.

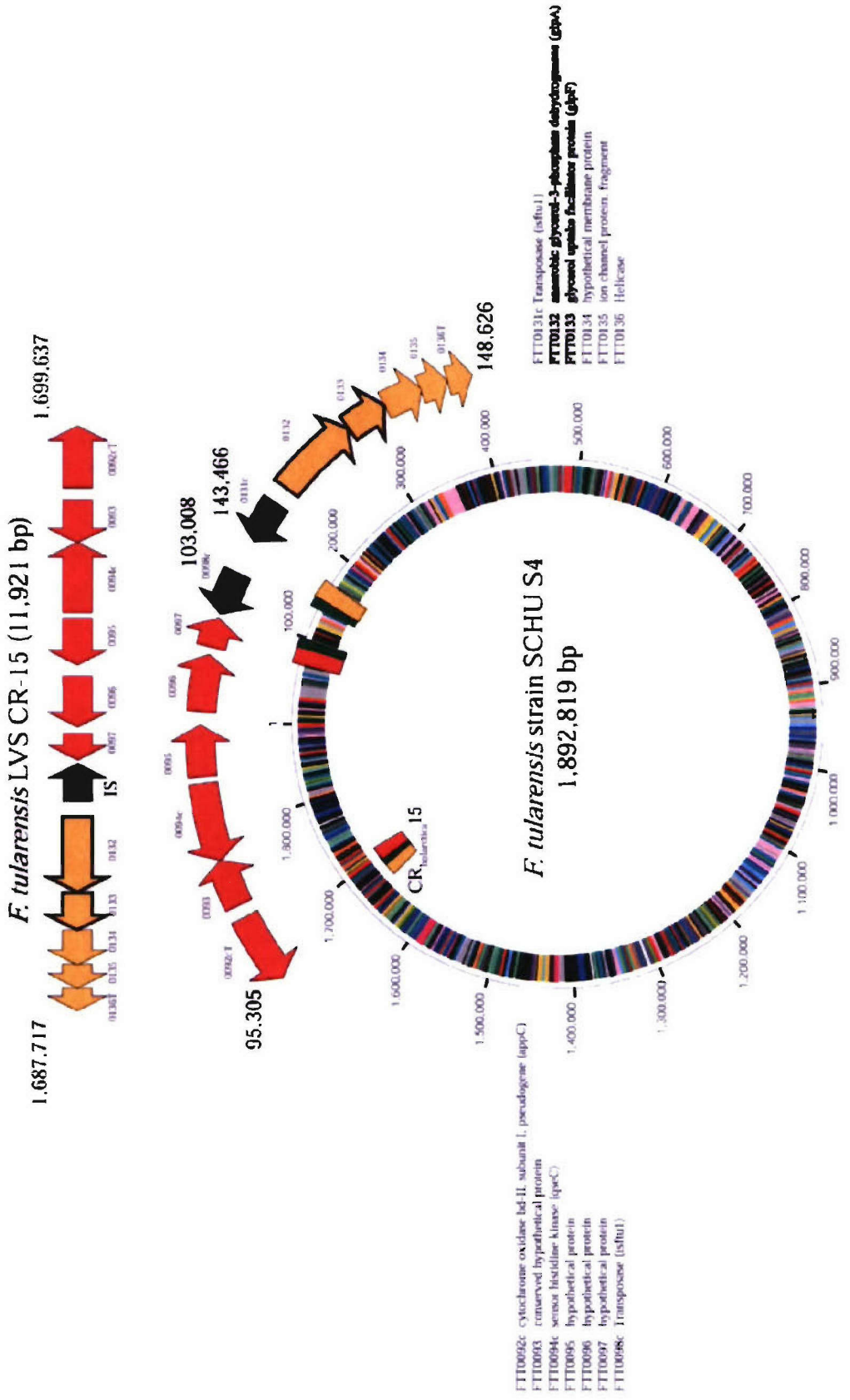


Figure 4-8.

Juxtaposed $CR_{holarctica}$ Segments in Subsp. *tularensis*.

The figure shows juxtaposed $CR_{holarctica}$ segments from CR1 and CR16, CR4 and CR10, and CR13 and CR15 comprising contiguous segments in SCHU S4 in panels 4-8a, 4-8b, and 4-8c, respectively. Genomic content is conserved between genomic segments bearing the same color and connected by small arrows. Crossed small arrows show inverted syntenic regions between the two respective genomes whereas parallel small arrows show the same synteny. The yellow line with two black wavy lines in between each LVS CR represents the large span of genome sequence separating the respective CR.

Figure 4-8:

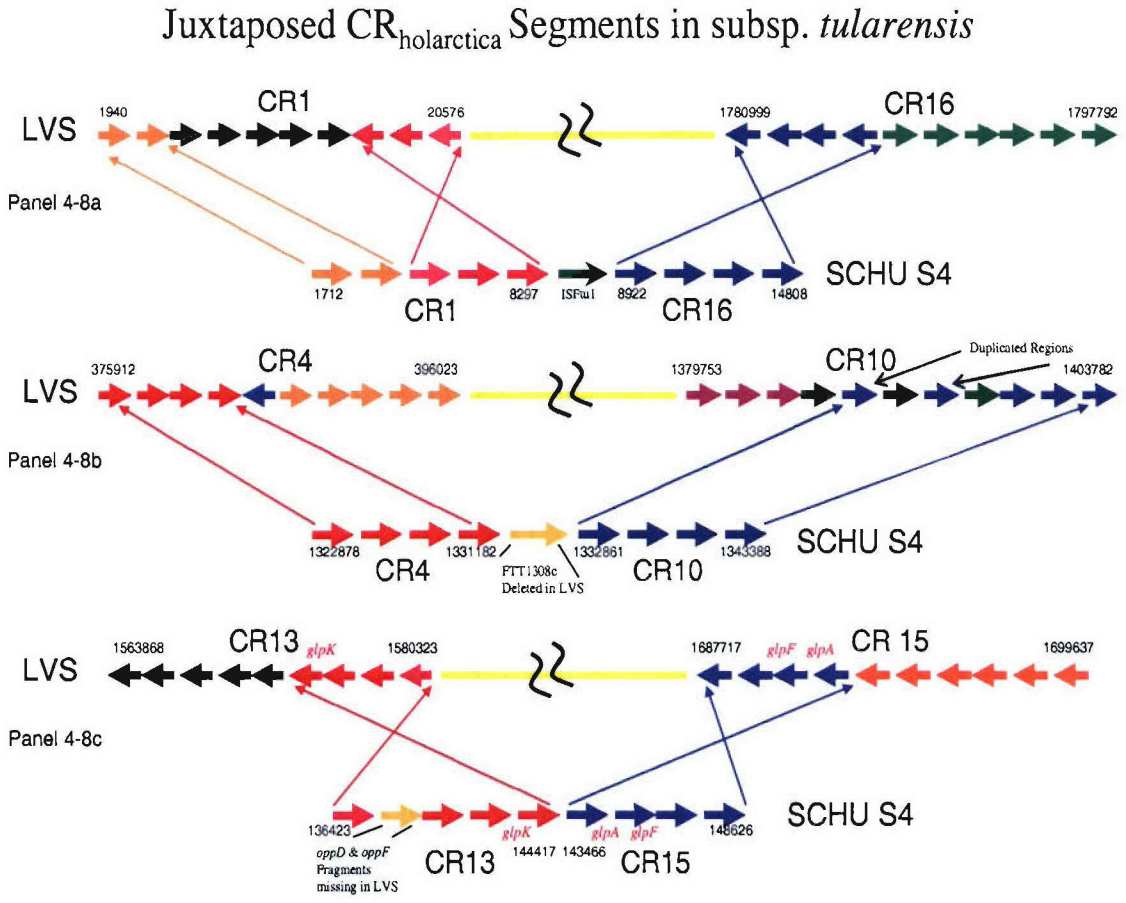
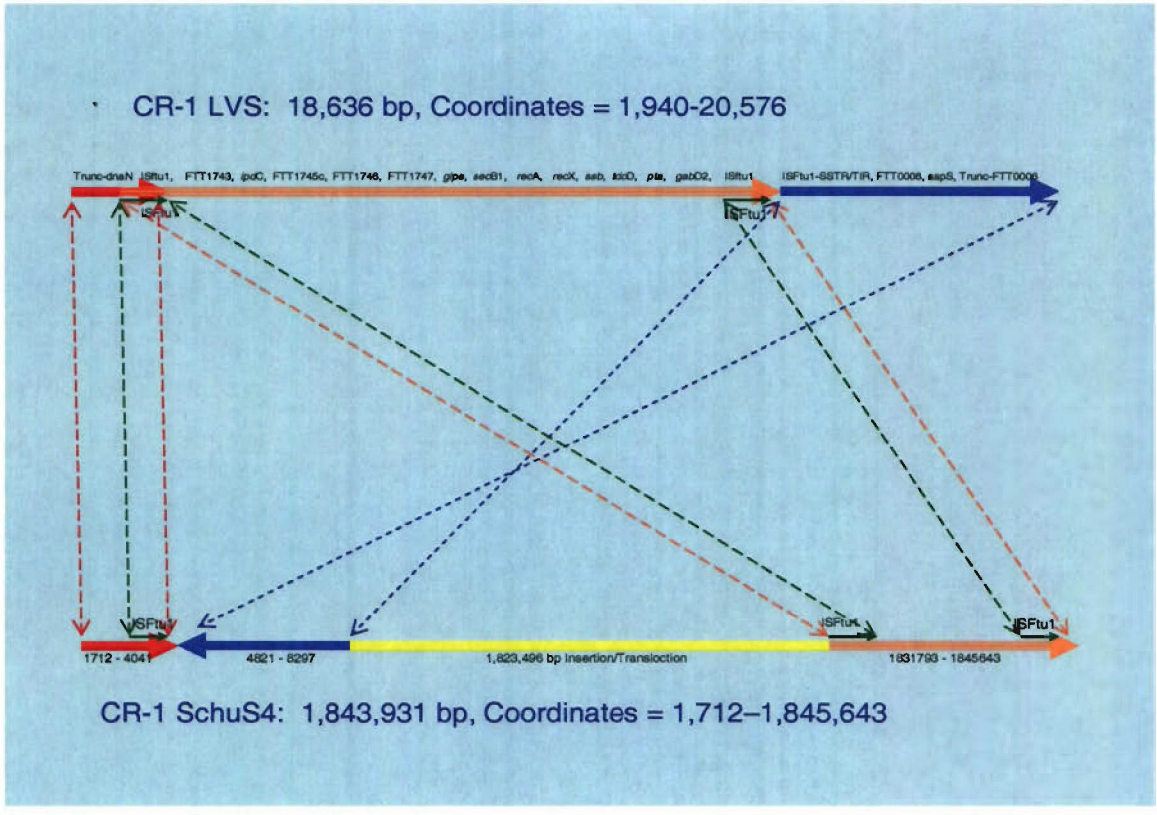

Juxtaposed CR holarctica Segments in subsp. tularensis
LVS
1940
CR1
20576
1780999
CR16
1797792
Panel 4-8a
1712
CR1
8297
ISFtu1
8922
CR16
14808
SCHU S4
LVS
375912
CR4
396023
1379753
CR10
Duplicated Regions
1403782
Panel 4-8b
1322878
CR4
1331182
FTT1308c
Deleted in LVS
1332861
CR10
1343388
SCHU S4
LVS
1563868
CR13
glpK
1580323
1687717
glpF
glpA
CR 15
1699637
Panel 4-8c
136423
oppD & oppF
Fragments
missing in LVS
CR13
glpK
144417
143466
glpA
glpF
CR15
148626
SCHU S4

Figure 4-9.

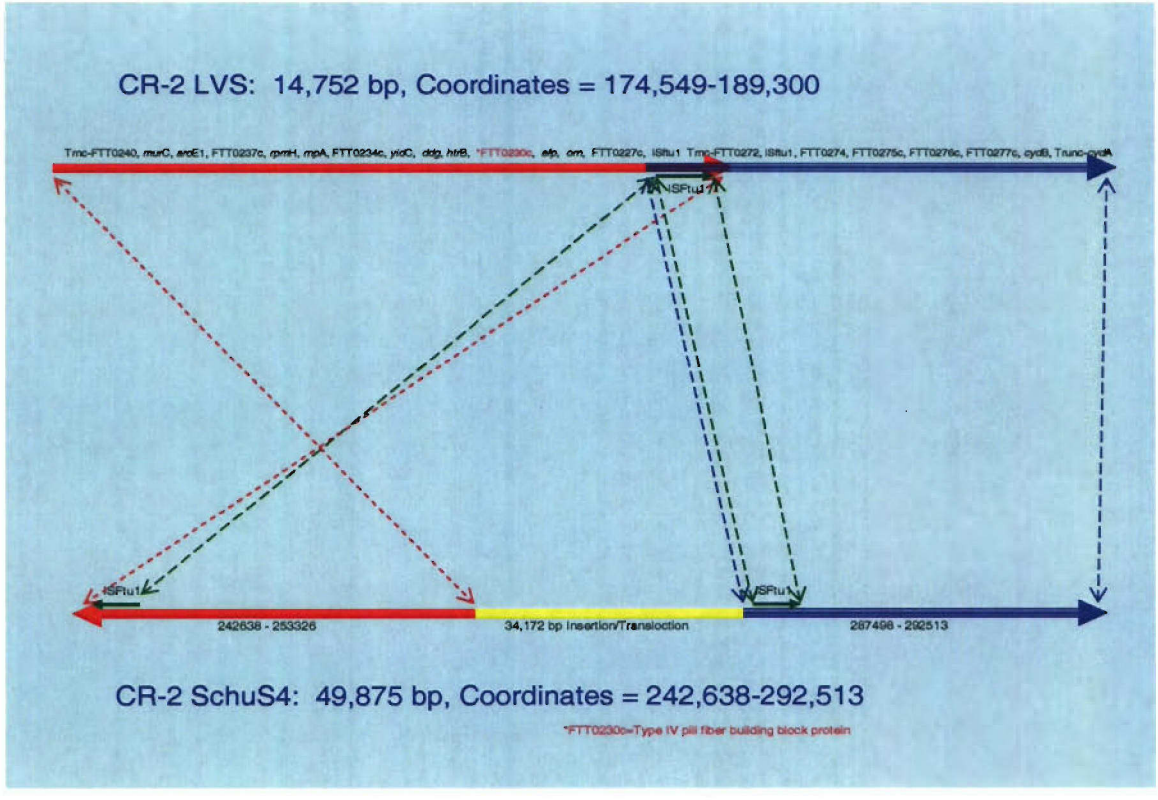
Individual $CR_{holarctica}$ Mapped onto Linear *F. tularensis* SCHU S4 Genomic Regions.

Each $CR_{holarctica}$ for CR1-CR17 is shown mapped individually onto the corresponding SCHU S4 genome in panels 4-6a-q, respectively. Upper line drawings represent the indicated $CR_{holarctica}$ LVS-specific genome content between its coordinates, whereas the lower drawing indicates the corresponding structure between the least and greatest SCHU S4 coordinates. Same-colored genome segrments in both LVS and SCHU S4 indicated homologous genomic content. Parallel-dashed lines represent segments in both genomes with the same synteny, whereas crossed-dashed lines represent inverted genome segments. Gene names or locus tags for all genes within each genomic subsegment are above their corresponding position within the LVS-specific $CR_{holarctica}$ subsegment. Gene names or locus tags designated by a prefix or suffix "T" indicate genes truncated due to the beginning or ending of the $CR_{holarctica}$ clone, or due to altered arrangements of genomic structure or content within SCHU S4 as compared with LVS. Genes, or their corresponding locus tags, of virulence- or biochemically-significant function are highlighted in red and defined below each SCHU S4 CR line drawing. Yellow lines in SCHU S4 segments represent insertions of SCHU S4-specific regions with no corresponding LVS sequence for the indicated CR. Rare hatched lines in LVS segments represent regions where no corresponding sequence exists in SCHU S4.

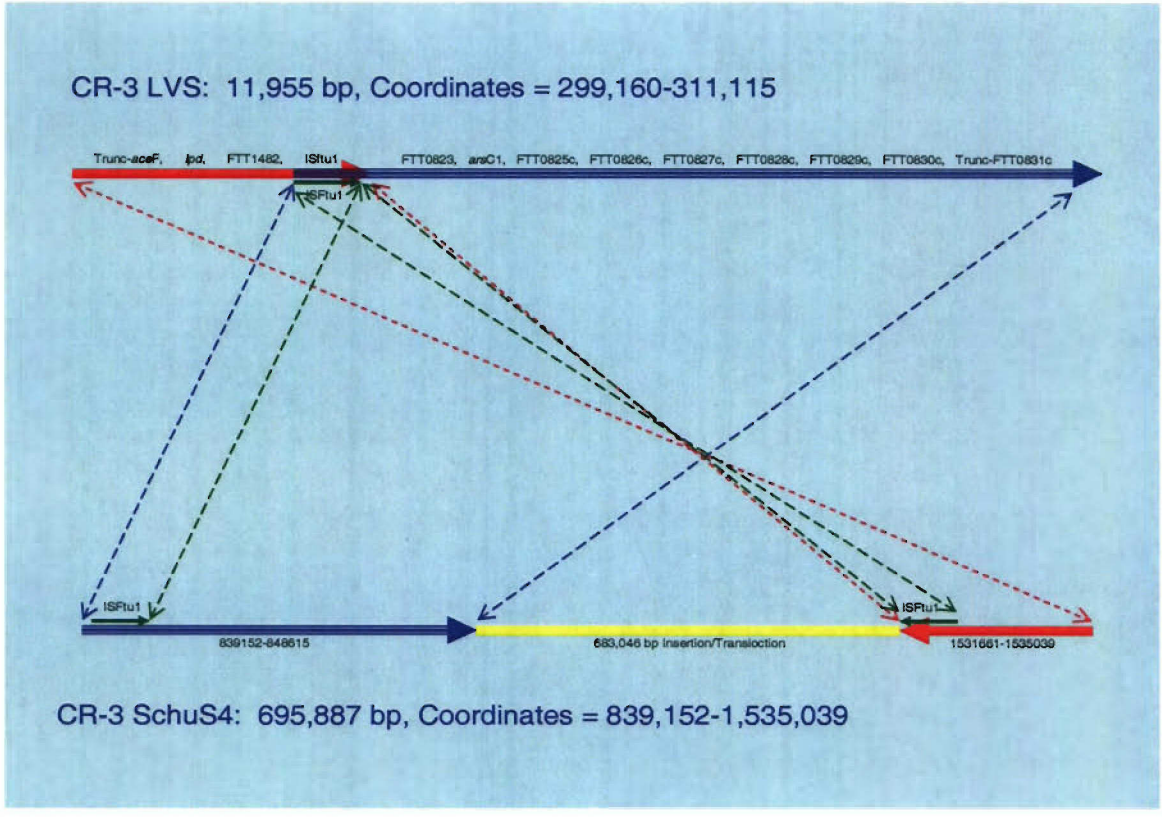
Panel 4-9a:



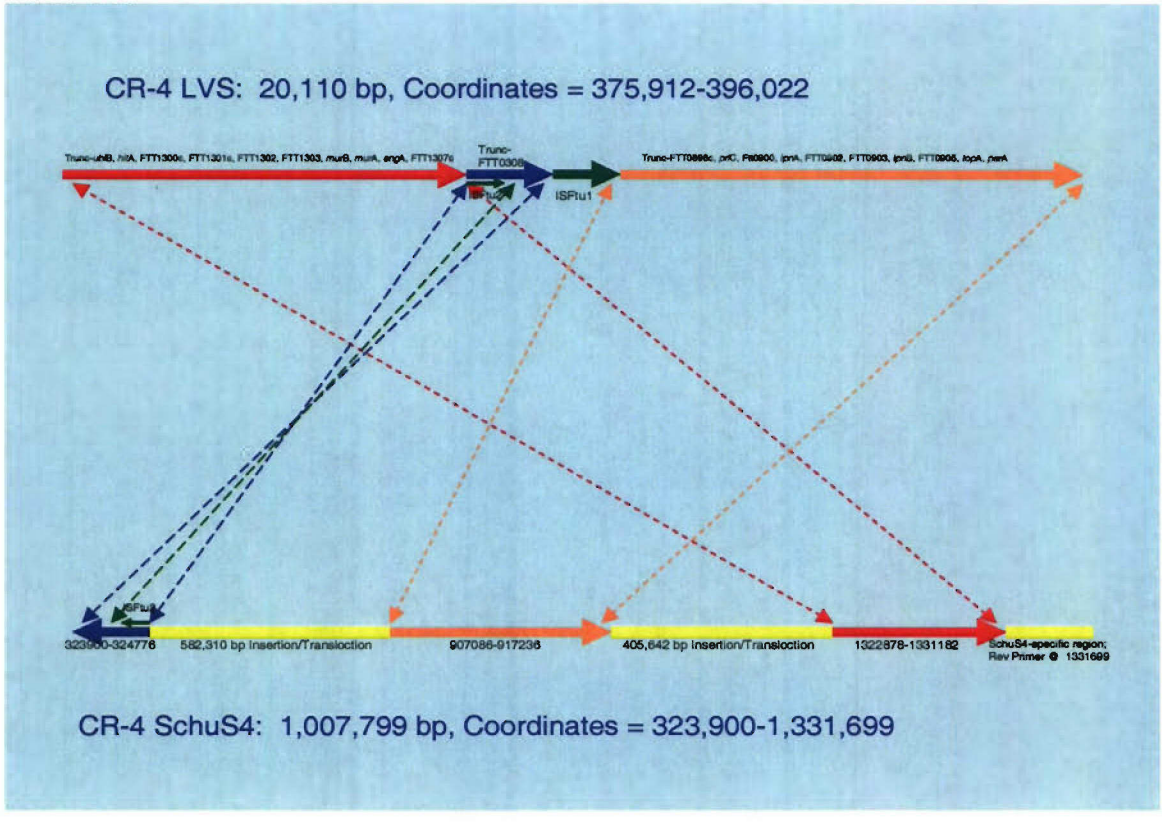
Panel 4-9b:



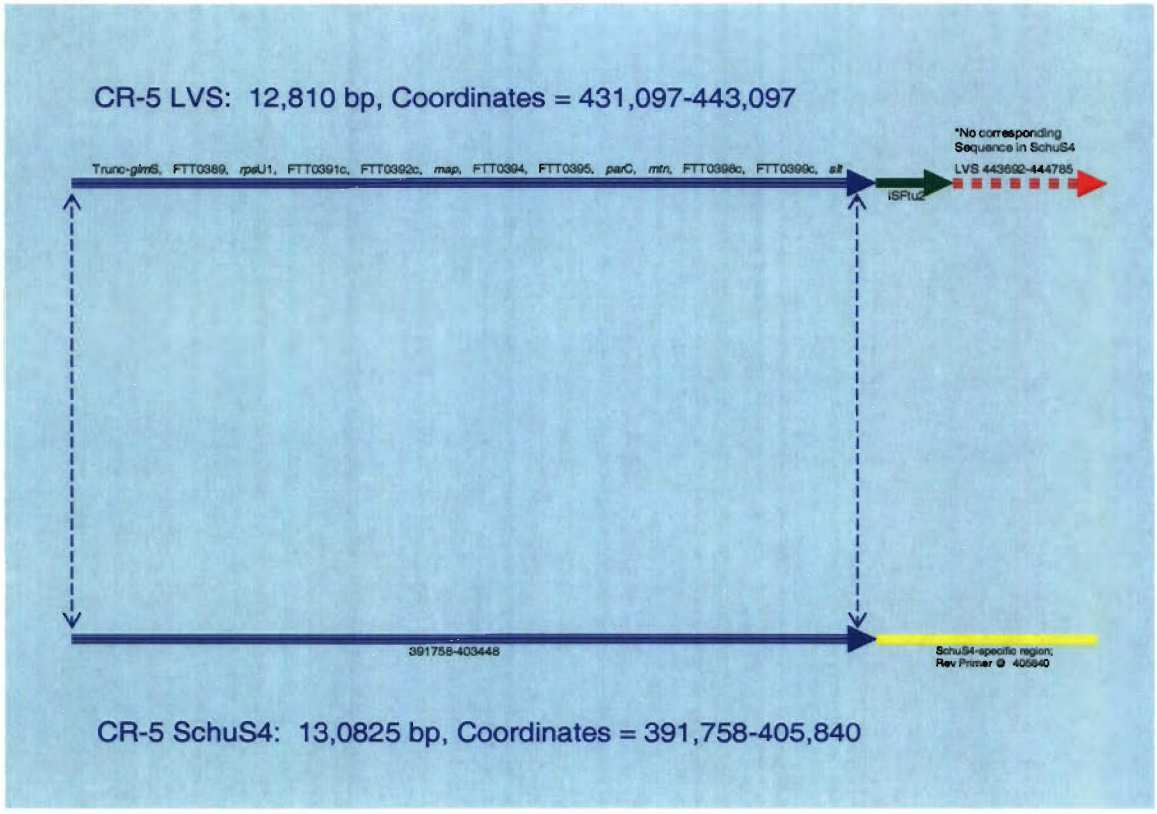
Panel 4-9c:



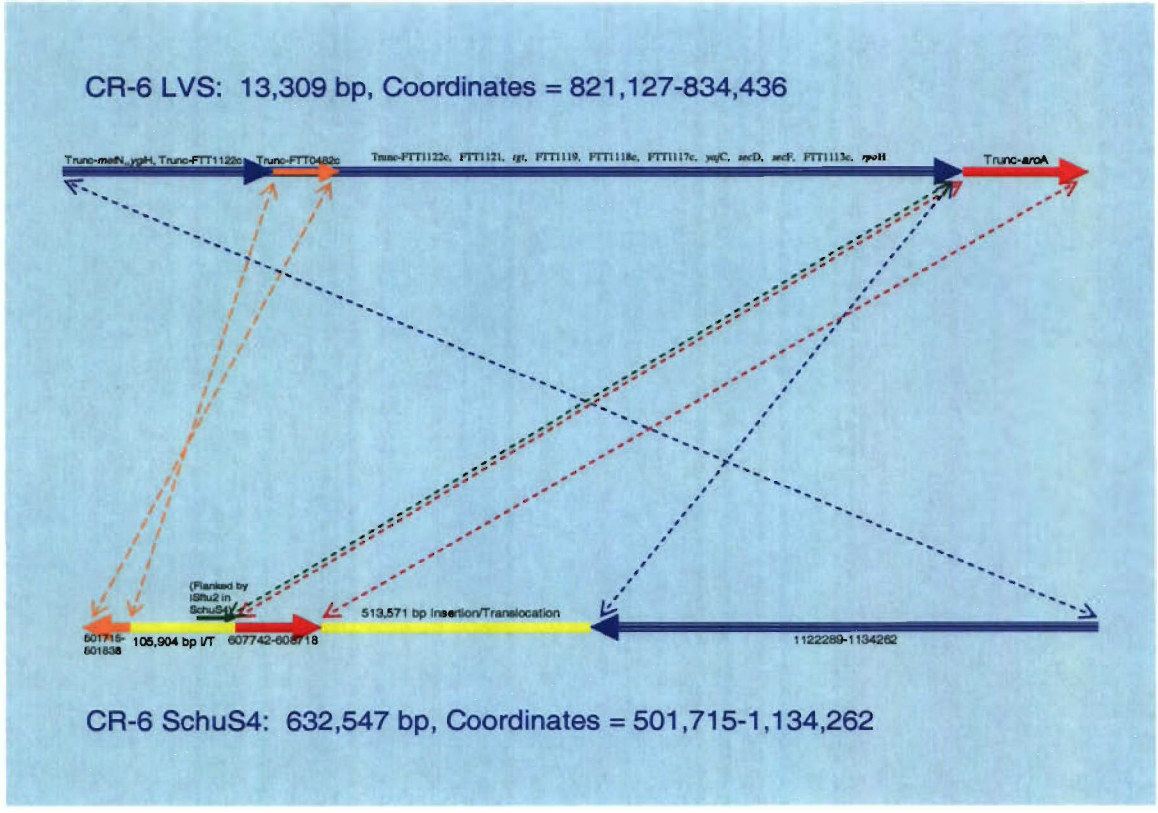
Panel 4-9d:



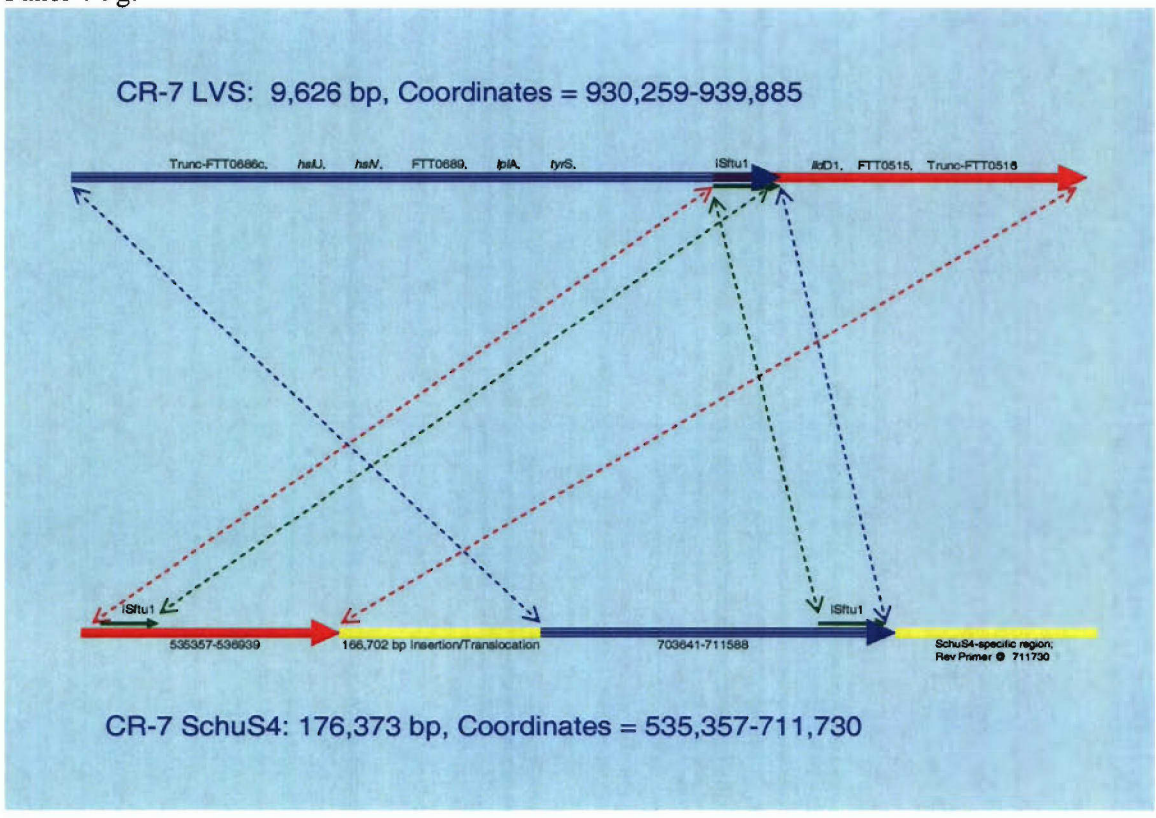
Panel 4-9e:



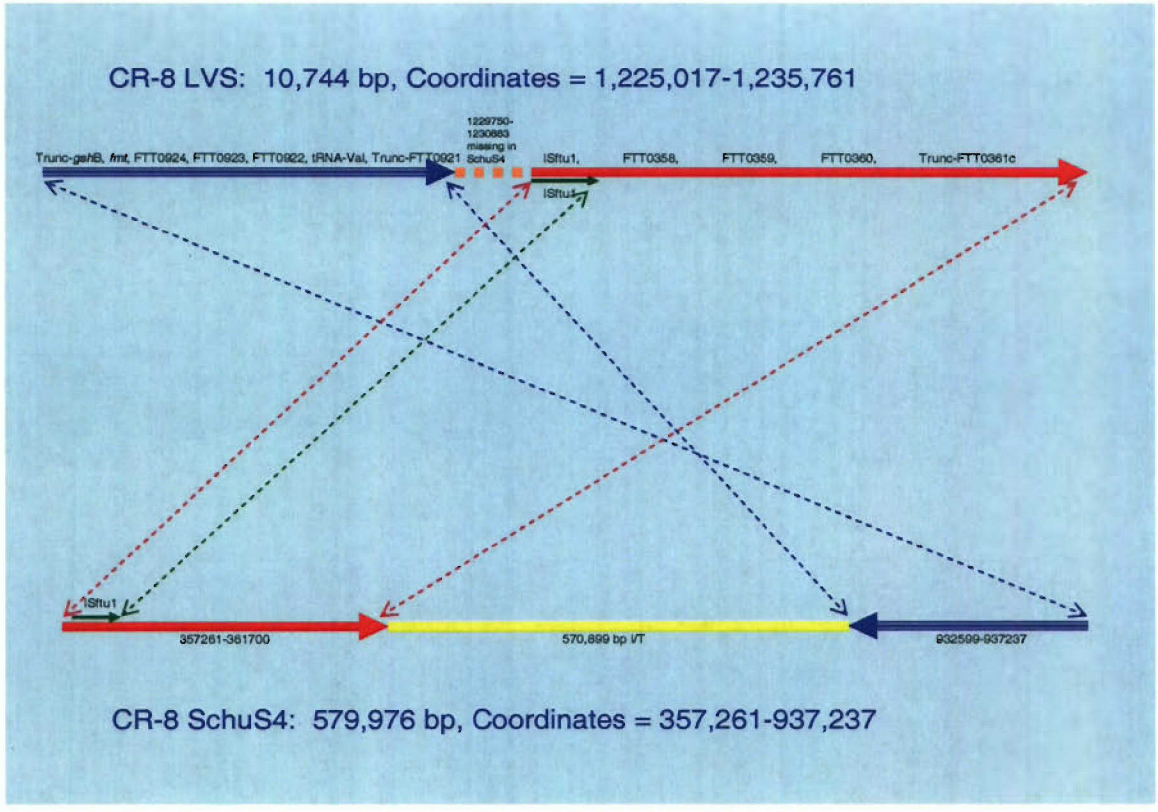
Panel 4-9f:



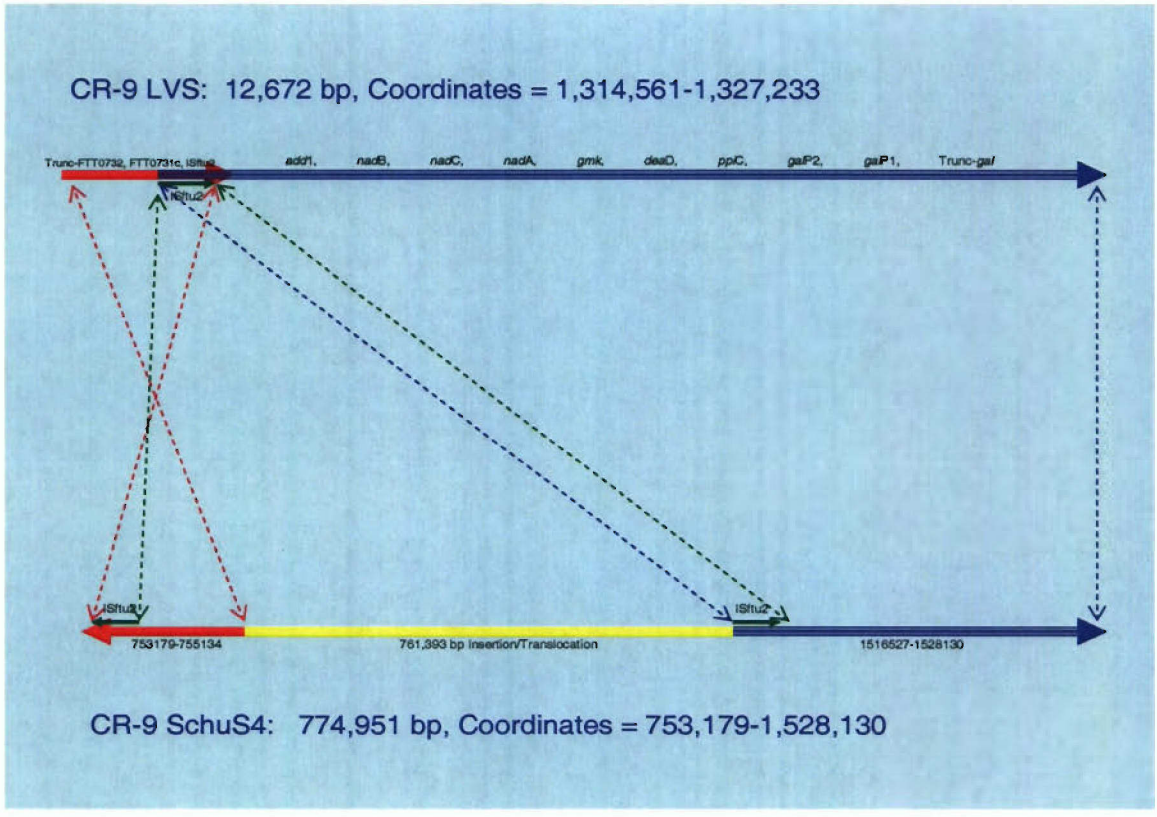
Panel 4-9g:



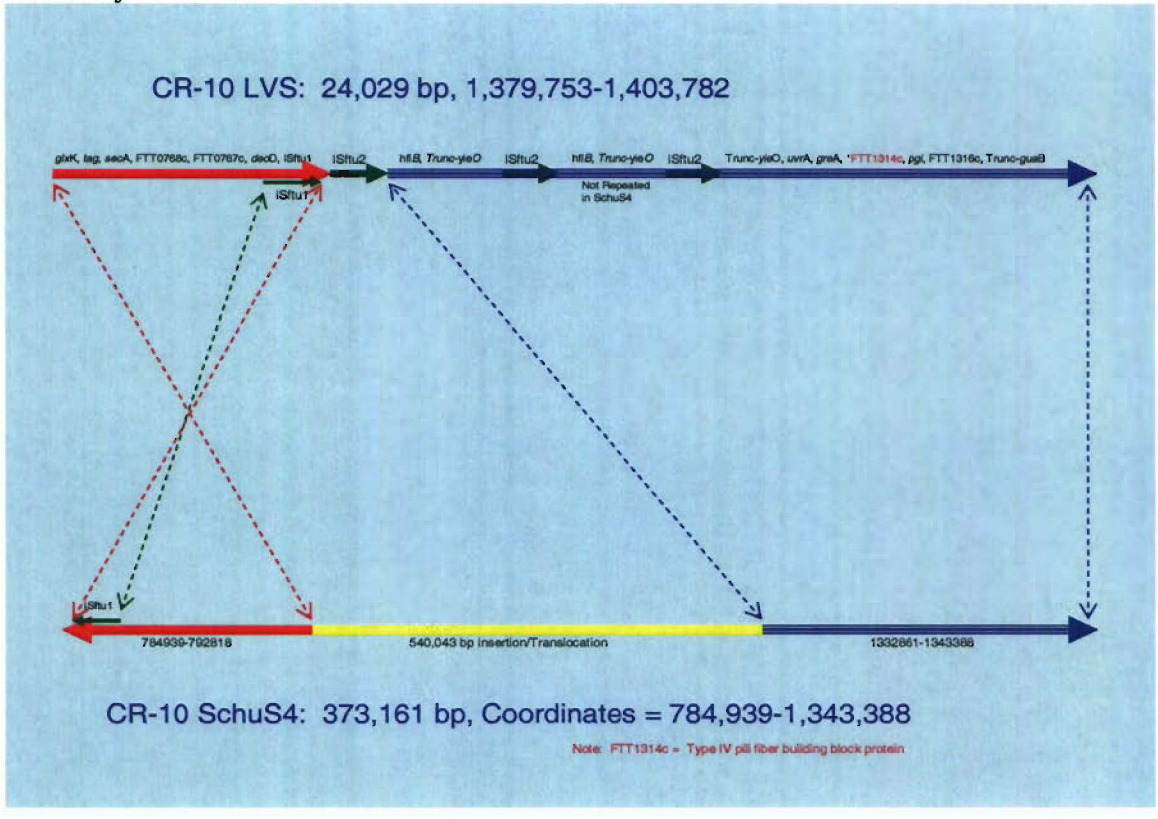
Panel 4-9h:



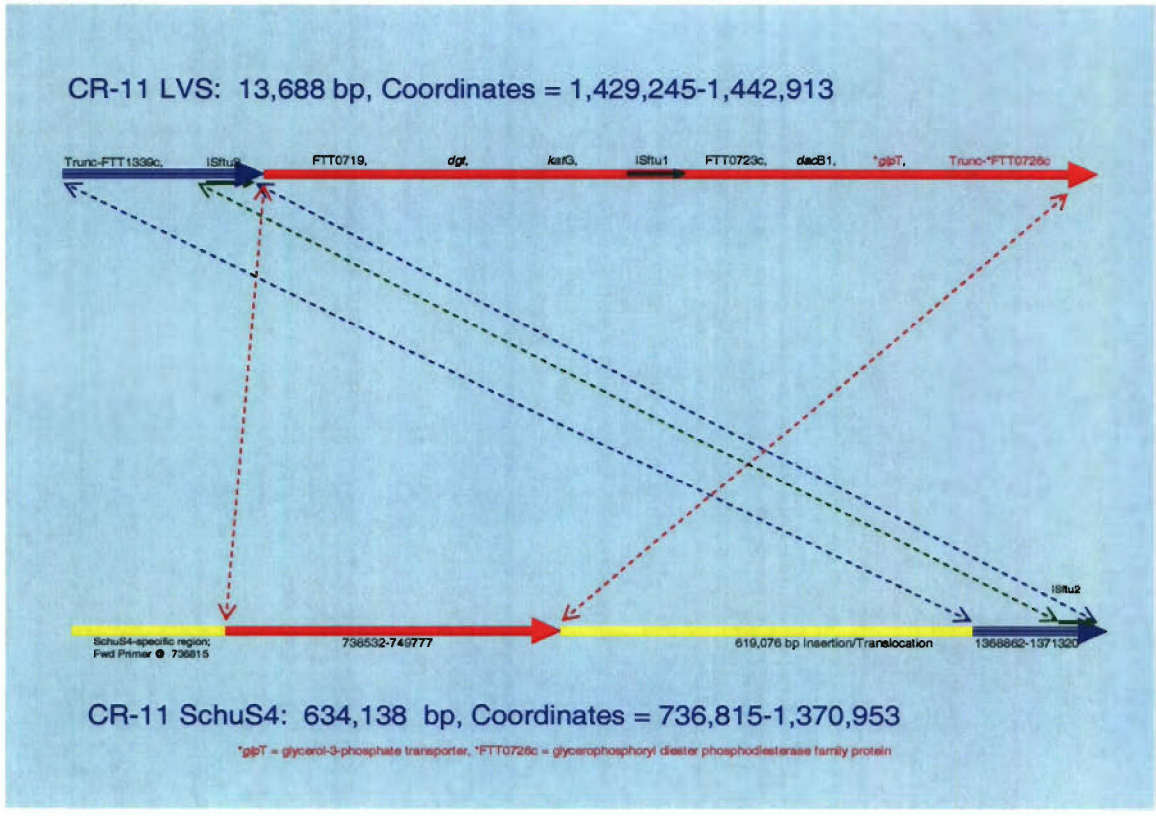
Panel 4-9i:



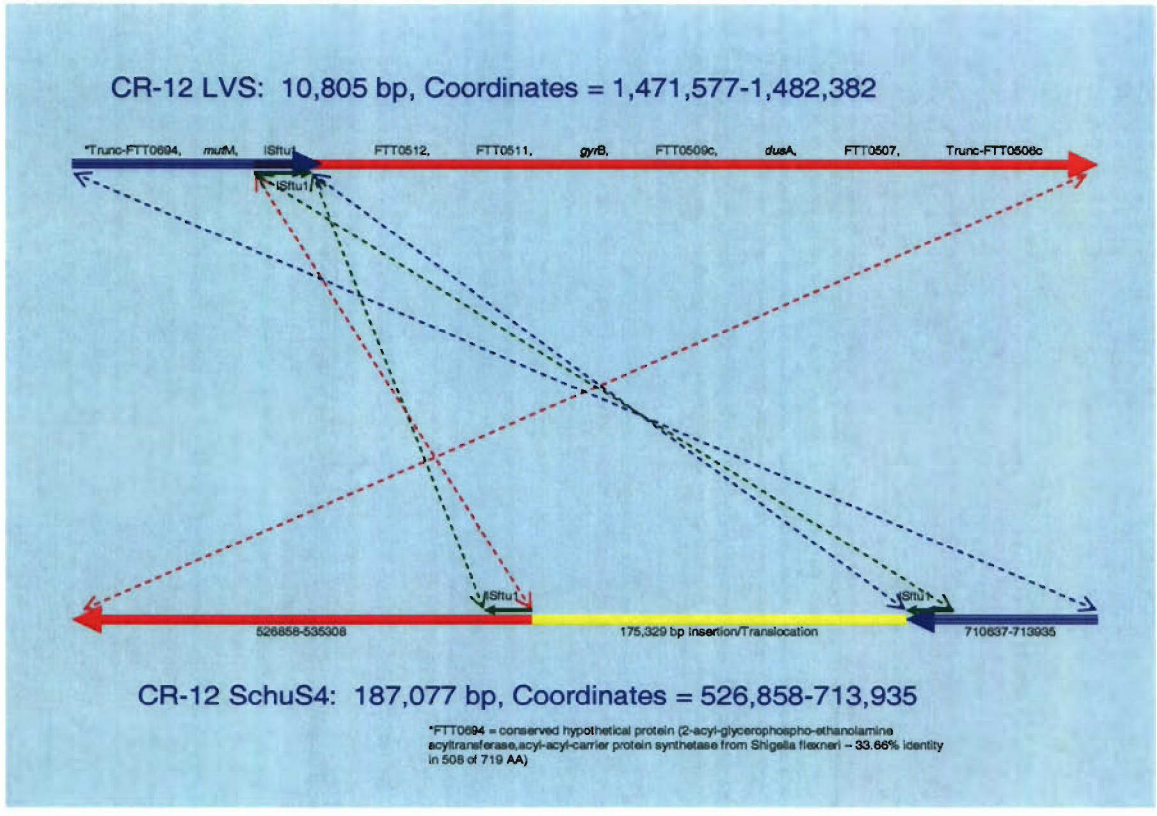
Panel 4-9j:



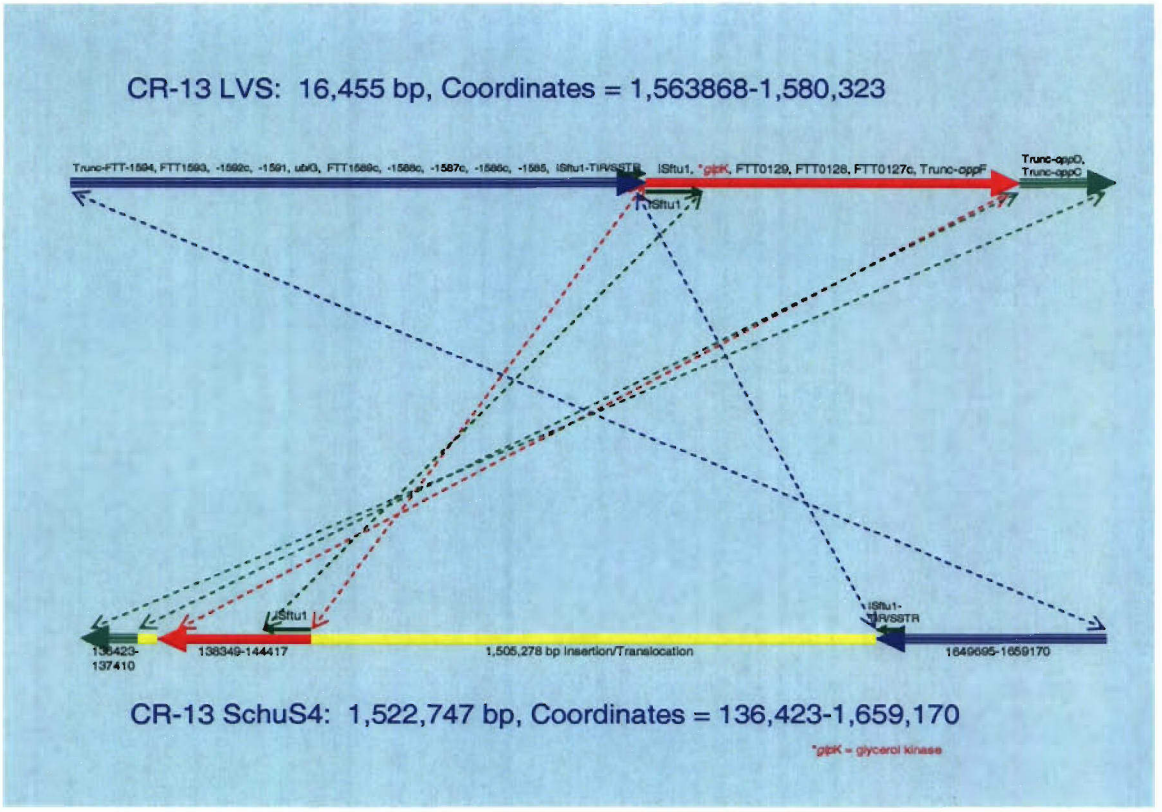
Panel 4-9k:



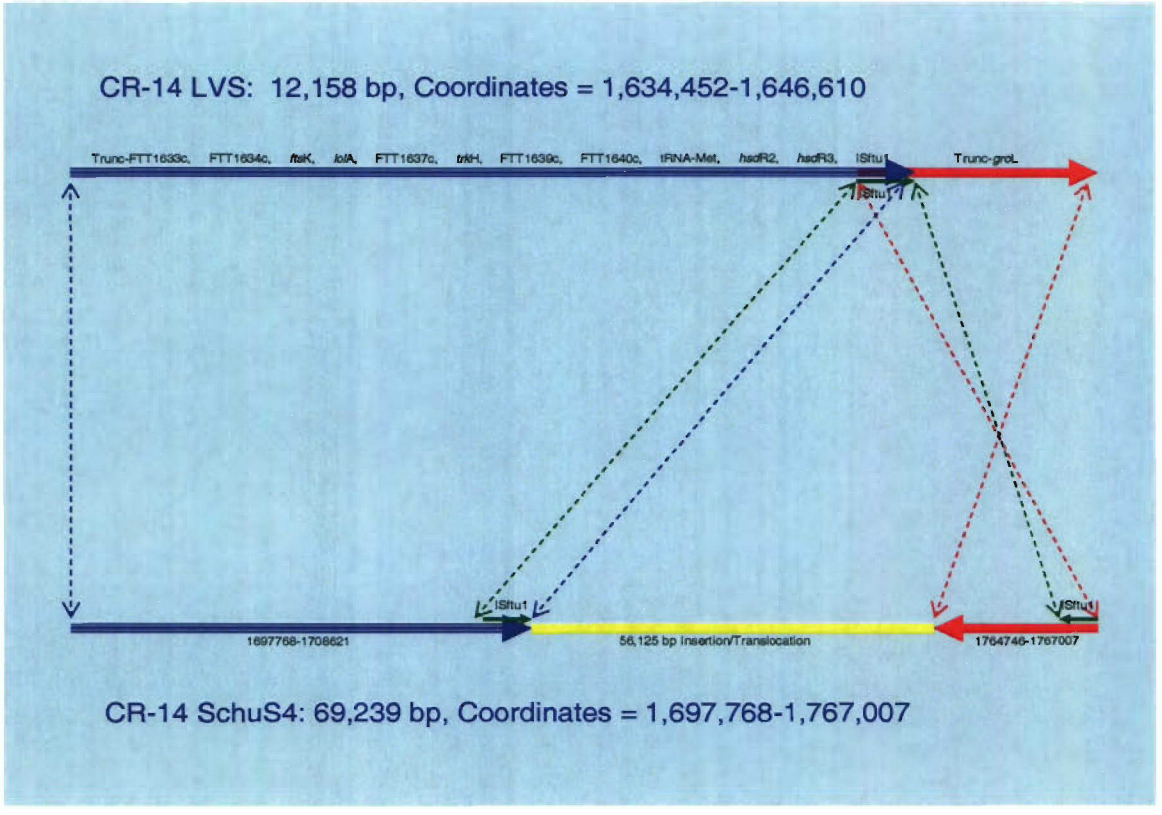
Panel 4-9l:



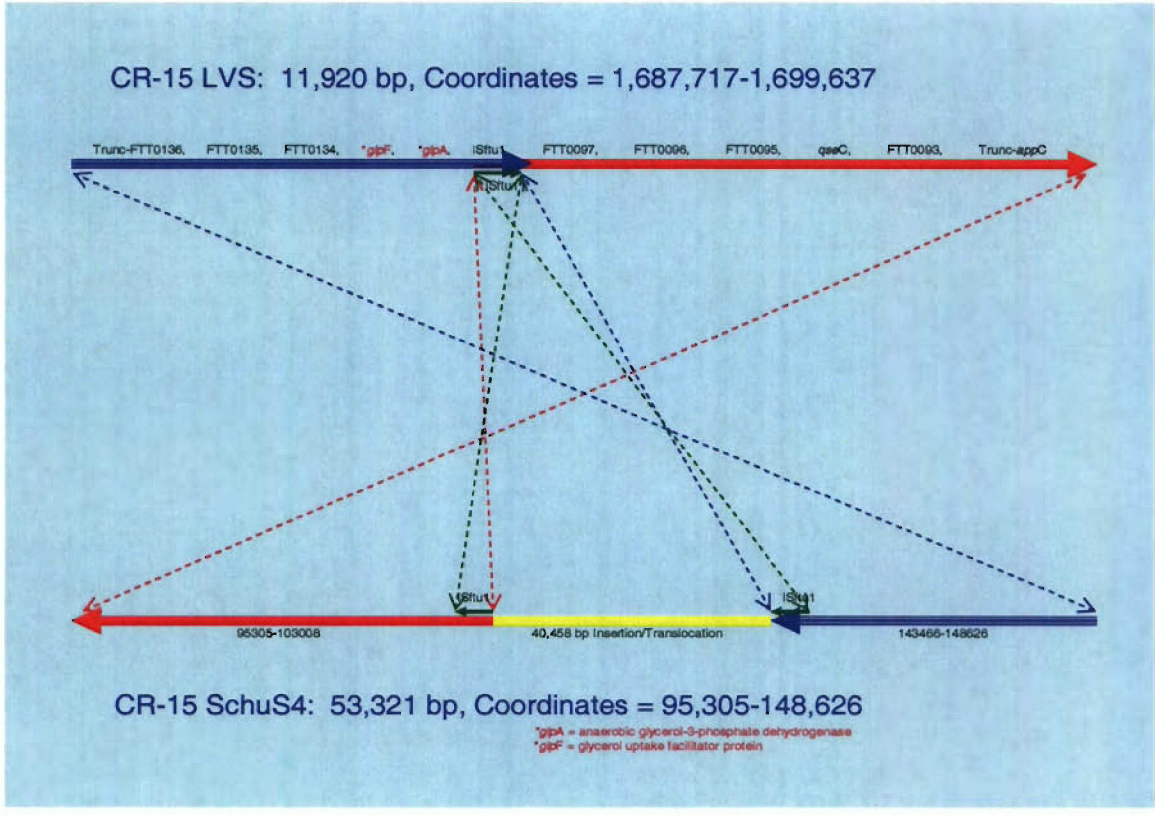
Panel 4-9m:



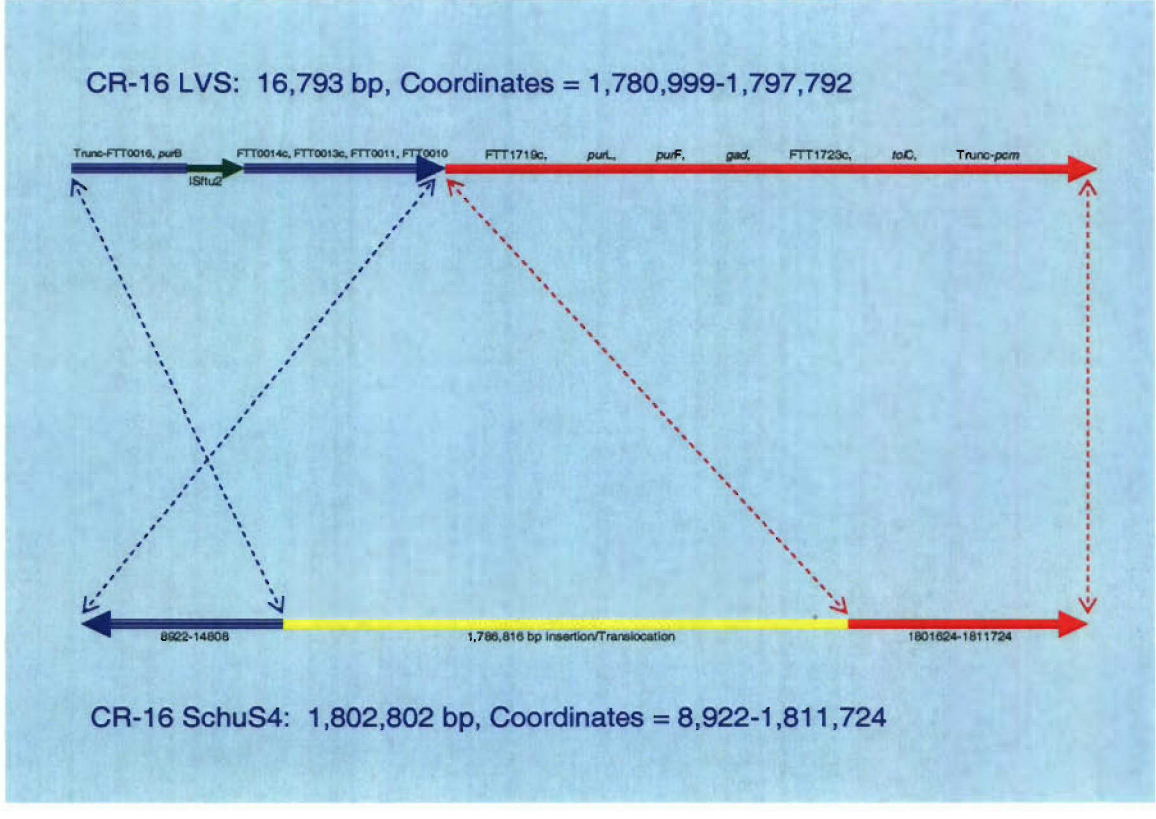
Panel 4-9n:



Panel 4-9o:

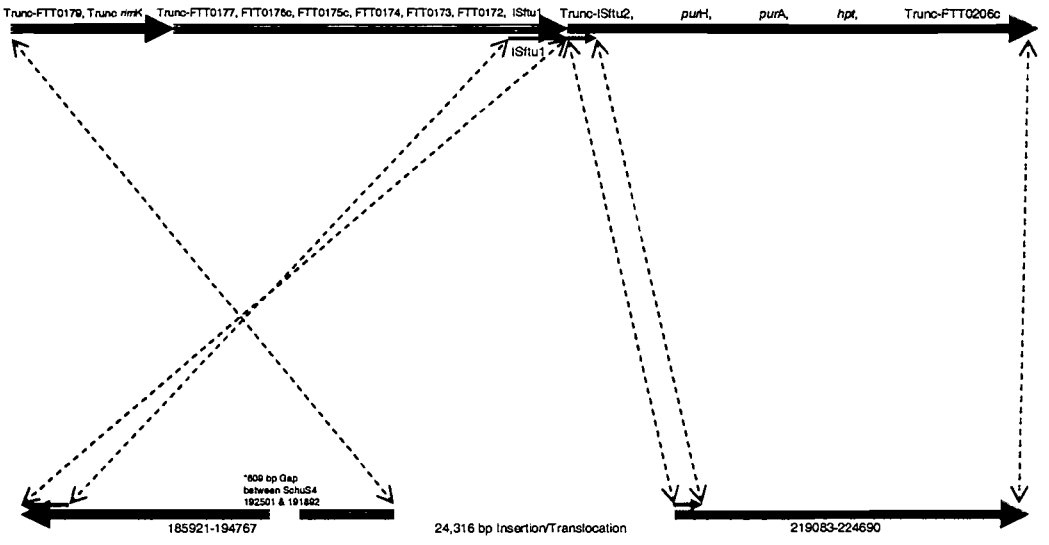


Panel 4-9p:



Panel 4-9q:

CR-17 LVS: 13,791 bp, Coordinates = 1,850,023-1,863,814



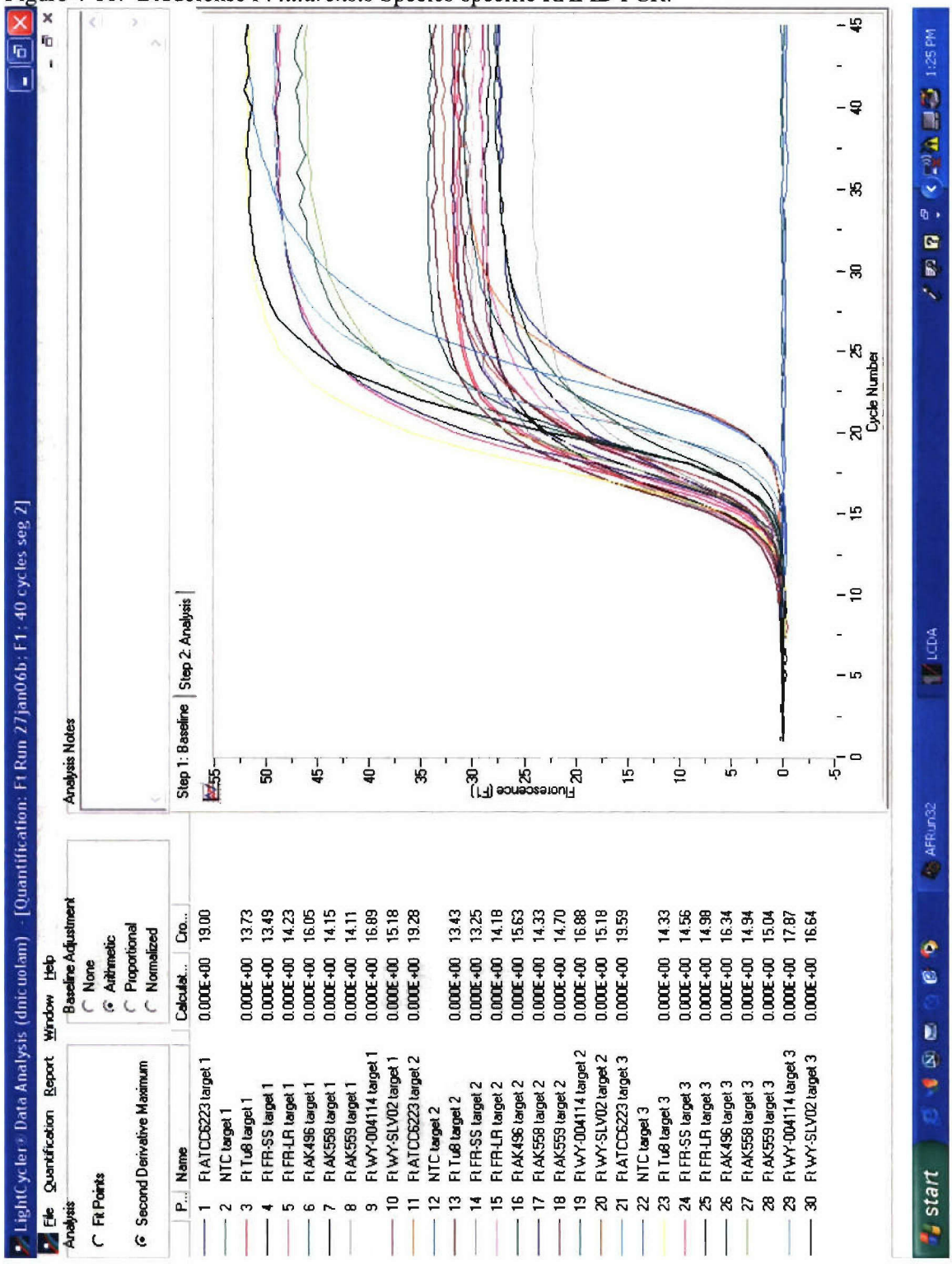
CR-17 SchuS4: 38,769 bp, Coordinates = 185,921-224,690

Figure 4-10.

Biodefense *F. tularensis* Species-specific RAPID PCR.

Shown in the figure are results from the RAPID instrument after running all three of the AFIP's biodefense *F. tularensis* species-specific PCR targets. As shown, all samples, including the strains from Alaska (AK), France (FR), a representative strain from Spain (Tu-8), the interesting Wyoming CR10-A3 strains, and the ATCC-6223 strain are positive for targets one, two, and three, respectively.

Figure 4-10: Biodefense *F. tularensis* Species-specific RAPID PCR.



CHAPTER 5:

Conclusions and Future Objectives

5.0 Overview

The results of the experiments in this dissertation support the hypothesis that numerous IS elements contained within the *F. tularensis* genome are responsible for the observed geographic-, subspecies-, and strain-associated genotypic differences as well as some of the associated biochemical and virulence phenotypic differences. This chapter will first discuss the significance of the RD_{Spain} polymorphism (from chapter 3), and in particular its potentially higher virulence in humans than other subsp. *holarctica* strains. Subsequently, PESM results (from chapter 4), will be discussed, especially the extensive IS element-mediated rearrangements evident between subsps. *holarctica* strains and *tularensis* strains which appear to have driven some of the subspecies-specific genomic and biochemical differences observed. In addition, the $CR_{holarctica}10$ genotypes seen among the subsp. *tularensis* strains tested as well as the overall uniqueness of two of the strains tested will be further discussed. Following discussion of PESM significance, the chapter will review mechanisms of *F. tularensis* divergence, both molecularly and geographically. This section will present models of *F. tularensis* subspecies genetic divergence including a new model accounting for *de novo* subtypes and/or taxonomic units based on collective observations from chapters 3 and 4. Also, natural mechanisms for geographical divergence of the *F. tularensis* subspecies will be discussed. The fifth section of the chapter will provide a novel PCR-based algorithm for differentiating and genotyping among the subspecies and strains of *F. tularensis*. In the final section I will discuss our results in context with current *F. tularensis* research efforts as well as propose future research including how our tools and advanced understanding of *F. tularensis* may be applied to the better understanding other microbial organisms.

5.1 Significance of the RD_{Spain} Polymorphism:

Chapter 3 described the first ever identification of an *F. tularensis* subspec. *holarctica* with a distinct polymorphism referred to as region of difference $(RD)_{Spain}$. The results of this study demonstrate that strains carrying RD_{Spain} are restricted to geographical regions within Spain and France, and are easily identifiable by PCR. While MLVA and other genotyping methods also provide some support for geographic differentiation of these polymorphic strains including strains from the AFIP laboratory and Keim Genetics laboratory at Northern Arizona University (or NAU laboratory), the populations are not entirely resolvable into geographic clusters by such methods as 16S rDNA sequencing, AFLP, PFGE, REP-PCR, ERIC-PCR, RAPD-PCR, and MLVA due to lack of a single common molecular signature [2, 32, 33]. Our studies using DNA microarray analysis showed that phylogeographic variation can be detected at the whole genome level without sequencing, and the variation is concordant with phylogenetic analysis. In conjunction with epidemiological data, we believe the variation that was observed in our study is a consequence of recent epidemic spread of a highly-related clonal population, which apparently has undergone additional divergence and diversification as evidenced by the MLVA genotypes and subtypes now observed.

RD_{Spain} appears to be associated with increased virulence in humans. Given the aggressive nature of the Spanish outbreak, and even though no genes associated with virulence were discovered in the RD_{Spain} sequence coding region, I hypothesized that RD_{Spain} may confer, or is associated with, a more virulent subsp. *holarctica* phenotype. I further hypothesized that this strain may be more geographically diverse and not just limited to Spain, and perhaps responsible for many of the tularemia outbreaks which occur quite frequently in Continental Europe. To test this hypothesis, I gathered epidemiology data from literature searches of tularemia outbreaks in Europe and contacted numerous authors reporting such outbreaks in hopes of acquiring either isolates or DNA to test for the presence of the RD_{Spain} polymorphism. These efforts, however, were unsuccessful until I contacted Dr. Christine Lion regarding her report of a 2001 case

involving a rare subsp. *holarctica*-associated bacteremia in a non-immunocompromised human 56-year-old male [101]. The resultant collaboration resulted in successful transfer of two isolates (the one from that case report, FR-SS, and another from a 1993 human ulceroglandular case, FR-LR) to the AFIP which were included in our RD_{Spain} study. As shown in chapter 3, these new strains as well as two other AFIP French DNA samples (which originally matched the AFLP profile of the Spanish outbreak samples [33]) were tested and found positive for RD_{Spain}. Since testing of the large global *F. tularensis* strain collection (n=319 strains) from the Keim Genetics Laboratory, with only one exception, confirmed that RD_{Spain} isolates were found only from France and Spain (with the one exception being an unexplained isolate from Sweden). Collectively these data demonstrate that my initial hypothesis that RD_{Spain} was associated with widespread continental European outbreaks was invalid; but as explained in chapter 3, its association with an apparent increased virulence on the Iberian Peninsula is of significant interest. This observation must now be validated through functional genomic studies based on the WT RD_{Spain} strains which can be used to create knockout mutants for testing in an animal model.

The observation of multiple clonal subpopulations carrying RD_{Spain}, as determined by MLVA, suggests that subsequent divergence has most likely occurred following emergence of the original primary RD_{Spain}-positive clone from a wild-type European subsp. *holarctica* strain. The IS-mediated mutation responsible for RD_{Spain} was most likely an infrequent event due to the relative stability of the IS elements in *F. tularensis* as suggested by Thomas *et al.* [26], as compared with relatively higher mutation rates known to occur at VNTR loci [73].

While it may be possible that the RD_{Spain} deletion has occurred on two independent occasions, the hypothesis that it occurred only once seems more parsimonious and can be supported, in-part, by the fact that our global strain sets have revealed no RD_{Spain}-positive strains in the numerous New World strains tested. If indeed two separate RD_{Spain}-positive clones have convergently evolved, then it could suggest that a selective amplification of the RD_{Spain} genotype confers a selective functional advantage, such as the hypothesized virulence. While the MLVA

data presented in chapter 3 show only subtle differences in RD_{Spain} (3 bp and 6 bp differences at the M4 and M22 loci, respectively) which, for now, are suggestive of two main genotypes within the RD_{Spain}-positive strains, previous reports suggest that MLVA alone may not be sufficient to definitively detect the true phylogenetic relationships within the global subsp. *holarctica* strains since they have lower genetic diversity as compared with subsp. *tularensis* strains [2, 78]. Due to this decreased genetic diversity in *holarctica* strains, further studies are required to definitively differentiate between the two main RD_{Spain} genotypes presented here.

Since the RD_{Spain}-deleted DNA segment and its adjoining flanking region of one strain (Tu-19) from the AFIP subset has been sequenced, but none from the collection at the NAU laboratory have been sequenced, further isolate DNA sequencing may reveal differences regarding the nature of the IS elements and their adjacent direct repeats involved with the particular deletion within each laboratory's respective subset of strains. IS-mediated mutations occur through the recombination of direct repeats occurring within the genome. In general, direct-repeat deletions leave a single direct repeat after excision of the deleted segment. This direct repeat is a composite repeat formed by fusion of the left and right flanking repeats. Although the actual deletions appear identical within each subpopulation by RT-PCR, the actual DNA sequences associated with the direct repeats flanking the deleted region may be different and discernable by sequencing.

5.2 Significance of the PESM Experiments:

The work presented in chapter 4 complements our previous micorarray-based work [91] in providing a more comprehensive catalogue of genomic differences between the *tularensis* and *holarctica* subspecies of *F. tularensis*. The work itself involved several innovative approaches which are worthy of mentioning. First, our strategy of building the genomic library using a Lambda-Dash replacement vector system proved highly successful and overcame our initial limitation experienced when we attempted to construct it using a topoisomerase-mediated ligation

plasmid vector. The latter strategy did not tolerate the large inserts (~10 kb) initially attempted; and furthermore, use of the Lambda system likely provided a larger degree of representation within the library itself due to overcoming the toxicity commonly associated with plasmid-based strategies. One disadvantage to our strategy, however, was that the large size (average of ~14.2 kb) and variable size distribution of our clone inserts (hereafter called amplicons) necessitated that each amplicon be measured so as to accurately compare to its coordinate-based size; but employment of gel-electrophoresis, an appropriate size standard, and the Syngene GeneTools analysis software proved successful in providing the actual amplicon sizes. Recovery of the amplicons themselves initially proved challenging (from both a technical and time-management standpoint) by traditional methods of growing liquid Lambda cultures and subsequent Lambda-DNA extractions; but this challenge was overcome by *TaKaRa* long-range PCR amplification directly (or as defined in chapter 4, DPA) from the plaques.

Whereas the single IS element-mediated RD_{Spain} didn't provide direct genetic insight into a possible association with increased virulence, the high number of IS element-mediated rearrangements apparent from the PESM study provided a more complete explanation of potential virulence and biochemical differences between subsp. *tularensis* and *holarctica*. We described at least 17 substantial genomic events which have occurred during divergence of the *tularensis* and *holarctica* subspecies. These results are provided in detail in chapter 4 and include differences in location and/or organization due to rearrangements affecting such genes as those encoding Type IV pili and glycerol fermentation pathway enzymes. Also included from the analysis are truncations/interruptions of several genes in subsp. *holarctica* with respect to *tularensis*, including an AceF Pyruvate dehydrogenase E2-subunit of pyruvate dehydrogenase, a rimK 30S ribosomal protein S6 modification protein, and several proteins of unknown function, but which may be biologically important as pertains to metabolism or virulence.

Also significant from the PESM study was the observation that these genomic rearrangements appeared highly conserved across all the subsp. *holarctica* strains tested, in

comparison with demonstrably more heterogeneity among the subsp. *tularensis* strains tested. These results support the hypothesis that subsp. *holarctica* strains are more recently emerged from subsp. *tularensis* which is believed to be closest to the ancestral *F. tularensis* clone. In fact, among the *F. tularensis* strains tested, two predominate genotypes were detectable based on two distinct allele sizes at the CR10 locus. As described in chapter 4, we have designated these CR10 genotypes as A1 (~1200 bp) and A2 (~1500 bp). A third major genotype was present with a unique allele at CR10 as well as additional rearrangements at several other CR loci. We have designated this CR10 genotype as A3, and due to its uniqueness, it may in fact represent a new taxonomic unit as discussed below.

The two US strains which may represent a unique taxanomic unit within *Francisella* with the CR10-A3 genotype were originally recovered in Wyoming. Although our search using the PESM strategy was primarily focused on identifying population-specific regions of genomic difference, the genome organization observed in the subsp. *tularensis* strains WY00W4114 and WY-WSLVL02 is very intriguing. Their pattern of genome organization is clearly distinct from the *tularensis* and *holarctica* populations, sharing CR "alleles" at some loci with *tularensis* strains, CR "alleles" at other loci with *holarctica* strains, and unique alleles at still other loci. The diversity is such that we propose they represent a new taxonomic unit. Assuming that additional isolates can be found, I propose the name for this new subspecies to be *F. tularensis* subsp. *neotularensis* due to its *tularensis*-positive RD1-PCR result and predominance of *tularensis*-specific CR segments (MD-unpublished), *tularensis*-like PFGE and glycerol fermentation results, and the fact it is virulent in humans [91]. In addition, a detailed surveillance and epidemiology study is necessary, as it is for the RD_{Spain} holarctica genotype, to map their distribution and relative virulence in humans.

5.3 Molecular Models Of *F. tularensis* Subspecies Divergence:

Svensson *et al.* have recently proposed an evolutionary model for the subspecies of *F. tularensis* based on unidirectional deletions [14]. According to the authors, the use of RDs for phylogenetic analysis relies on an assumption that unidirectional deletion (not insertion) events eventually become fixed in bacterial populations [111, 112]. Nine RDs from the works of Samrakandi *et al.* [91], Broekhuijsen *et al.*[90], and some identified *de novo* in the Svensson study, were included in their phylogenetic analysis. The conclusion of their analysis proposed an evolution of *F. tularensis* where the highly virulent subsp. *tularensis* preceded the appearance of the less virulent subsp. *holarctica.*

In the Svensson model, a common ancestor to all the *F. tularensis* subspecies was proposed containing intact genomic segments of all 9 RDs. *F. tularensis* subsp. tularensis appears to best represent this common ancestor by the presence of all nine RDs. Divergence from the common ancestor to subsp. *novicida* was proposed as resulting from extensive single nucleotide variations in seven genes as well as insertion of genomic content, apparently at all three subsegments of RD1: RD1a, RD1b, and RD1c. The subdivisions of RD1 in their study helped demonstrate the polymorphic nature of RD1 among the different subspecies; for example, RD1a is missing only from subsp. *holarctica-japonica* strains, whereas RD1b is missing only from N. American and Eurasian subsp. *holarctica* strains, and RD1c is missing only from subsp. *mediaasiatica* strains. The model next demonstrated the divergence of subsp. *mediaasiatica* from subsp. *tularensis* as having occurred by deletion of only one of the nine RDs, RD1c, and therefore appears genetically most like subsp. *tularensis* than the other subspecies. Divergence from subsp. *tularensis* to subsp. *holarctica-japonica* was proposed to have occurred by loss of four of the eight remaining RDs, and therefore subsp. *holarctica-japonica* appears to be an intermediate between subsp. *tularensis* and subsp. *holarctica* which has lost the remaining 3 RDs. The details of all RD deletions mentioned above, as well as a schematic diagram, are presented in the Svensson *et. al* article [14].

While agreeing in principle with the model presented above, the model must be updated to reflect our findings pertaining to *F. tularensis* subspecies divergence. As discussed above and in chapter 3, RD_{Spain}, while intact in all other subspecies, it has been found to be deleted exclusively from *F. tularensis* subsp. *holarctica* strains found in NW Spain and France, with one exception being from a single isolate in Sweden. In addition, the PESM PCR experiments discussed above and in chapter 4 have shown that the $CR_{holarctica}10$ (or CR10) PCR assay has successfully differentiated 3 different genotypes of strains shown to be subspec. *tularensis* by RD1 PCR. As previously mentioned, the genotypes at the CR10 locus are designated as A1, A2, and A3. Results from testing SCHU S4, which has been classified as an A.I. MLVA genotype [2, 78], likewise demonstrated that it has a CR10-A1 genotype. In addition, results from testing the *F. tularensis* ATCC 6223 strain, which has been classified as an A.II. MLVA genotype [2, 78], has likewise been shown to have a CR10-A2 genotype. For the purpose of further discussion, I conclude that a direct correlation exists between the CR10-A1 and MLVA-A.I. as well as between the CR10-A2 and MLVA-A.II genotypes. This correlation therefore allows the identification of a subspec. *tularensis* CR10-A1 or CR10-A2 allele to represent a MLVA-A.I. or MLVA-A.II. genotype, respectively, based on the single CR10 PCR locus. While these results correlating the CR10-A1/A2 alleles with the MLVA-A.I./A.II. genotypes appear straight forward, the results demonstrating the CR10-A3 allele were totally unexpected, and do not correlate with any known alternative subspec. *tularensis* (or any other *F. tularensis* subspecies, for that matter) genotype by MLVA or any other genotyping method. As demonstrated in chapter 4, this specific allele was present in only the two isolates of our strain collection, both from Wyoming, and which I have proposed the new subspecies name of subsp. *neotularensis* as previously discussed.

Based on the collective RD_{Spain} and CR10 experiments, the flowchart in figure 5-1 explains our proposed model for molecular divergence of *F. tularensis*. The schematic, while for the most part reflecting the one published in the Svensson *et al.* paper [14], differs by inclusion of deletion of RD_{Spain} as well inclusion of the polymorphic CR10 locus (plus additional CR

polymorphisms observed for subsps. *neotularensis*, *novicida,* and *japonica*) as mechanisms explaining the subsp. *holarctica*$_{Spain}$ genotype, the two subsp. *tularensis* genotypes, and the new subsp. *neotularensis*, respectively.

5.4 Natural Mechanisms of *F. tularensis* Subspecies Divergence:

In their article which was previously discussed in chapter 1, Farlow *et al.* concluded from their MLVA study [78] that a geographic correlation of distribution can be drawn from the subsp. *tularensis* A.I. and A.II. subpopulations. The A.I. subpopulation distribution is closely associated with the distribution of the tick vectors *Amblyomma americanum* (Lone Star tick) and *Dermacentor variabilis* (American dog tick). Both *D. variabilis* and the A.I. isolates occur primarily in central and eastern United States, (also known as the "human tularemia incident hotspot" or "lower Midwest tularemia focus", and including Missouri, Oklahoma, Kansas, and Arkansas), but also in California, Alaska, and British Columbia. The main geographic cluster of A.II. isolates appears associated with the distributions of 2 known tularemia vectors, *D. andersoni* (Rocky Mountain wood tick) and *Chrysops discalis* (deer fly), which have occurred primarily in the western United States as well as Ontario and Texas. Different rabbit hosts (*Sylvilagus* spp.=cottontail rabbits; *S. floridanus* for A.I., and *S. nuttallii* for A.II) as well as differences in elevation (higher for A.II.) were both implicated as factors associated with the distribution of the two A clades, with the current hypothesis being that A.I. may have served as a parental strain to A.II., with A.I. appearing more diverse and therefore older than A.II.

Currently, the literature doesn't support the idea that *F. tularensis* existed outside the United States prior to the early 20th Century. As for the relative age of *F. tularensis* within the United States, again the literature doesn't specify, but Farlow *et al.* [78] speculate that the. subsp. *tularensis* A.I. population may have been present as a robust population locally isolated within lower Midwest tularemia focus before colonization by European settlers who dispersed it throughout the continent. The advent of modern transportation and rabbit and hare exportation

practices for hunting helped distribute isolates of each respective clade away from their primary geographical focus. The single clade B from this study appears less diverse and more geographically ubiquitous in North America by comparison [78], and suggests a more recent divergence than the A.I or A.II. clades. Petersen and Schriefer provide an excellent review [17] of the primary discovery of *F. tularensis* in the U.S., as well as its apparent subsequent global mechanical spread. As reviewed [17], until 1925 it was widely believed that tularemia was limited to the United States. At that time in Japan, the similarity of tularemia with the hare disease, Yato-byo, was made and later confirmed as tularemia. Subsequent to that, *F. tularensis* was identified in the USSR in 1928 as the causative agent of "water-rat-trappers" disease. Soon thereafter, tularemia was reported also in Norway (1929), Canada (1930), Sweden (1931), and Austria (1935) [113].

The review by Petersen and Schriefer [17] also discusses that numerous animal hosts have been associated with tularemia, and it provides examples regarding several tularemia outbreaks. For example, as previously mentioned, the transmission to NW Spain appears to have been related to exportation of *F. tularensis*-infected hares from several countries, including France, in 1996. The emergence of tularemia in Kosovo in 2000l, however, appears linked to rodents from a waterborne source due to poor sanitation resulting from war. In the U.S. in 2002, an outbreak was traced to a Texas exotic pet facility involving domestic and international exportation of tularemia-infected prairie dogs, some of which were distributed to the Czech Republic and a Texas pet store. In these cases thus far, the subspecies isolated has be holarctica [17].

Regarding the first outbreak in Spain, based on the age of the earliest known RD_{Spain}-positive isolate contained in the strain set at the Keim Genetics Laboratory (as published in REF.[2]), our current understanding of the RD_{Spain} taxon is that it apparently first emerged (likely from a European WT subsp. *holarctica* strain) in France, as early as 1952, where it has remained. As late as 1996, the spread of a clone positive for RD_{Spain} appears to have occurred from France to

NW Spain through hares imported for hunting purposes (reviewed in REF.[17]) and was subsequently primarily for initiating of the 1997/1998 tularemia outbreak in that country. It remains unknown if RD_{Spain}-positive strains were involved in the second wave [35] since no strains were isolated for PCR analysis. As observed from the numerous human cases attributed to this variant subsp. *holarctica*, the disease in humans has been more severe than expected, and therefore expanded surveillance as proposed in chapter 3 is warranted until the suggested functional genomic studies can be performed to confirm the correlation.

Also discussed in the review was the sentinel *F. tularensis* subsp. *novicida* isolate from a human in Australia [17]. As one may presume that lagamorphs (Brown Hares) from England, first introduced in Australia in 1859 for hunting purposes, may have been linked as the natural reservoir of this particular isolate, no such association has been made. In fact, no occurrence of tularemia in the United Kingdom can be found in the literature. Based on the evidence thus far supporting a model of global divergence originating from the U.S. in the early 20th century, the hares from England originally introduced into Australia wouldn't yet have been infected. The fact that the Australian isolate was obtained following a waterborne exposure is in-line with our current understanding of subsp. *novicida*, but does suggest that an importation of some component of the aquatic reservoir associated with this subspecies may have occurred in Australia. Petersen and Schriefer offer no definitive explanation as to the presence of this isolate except to underscore the likelihood that *F. tularensis* may be more widespread than previously thought, and to raise the question of whether other *F. tularensis* subspecies are also present in Australia and elsewhere in the Southern Hemisphere [17]. A map showing the currently known global distribution of *F. tularensis* including *de novo* genotypes resulting from this dissertation is presented in figure 5-2.

5.5 PCR-based Genotype Differentiation Methods:

Currently our best resolving genotyping method is whole genomic sequencing followed next by CGH microarrays and PESM. Due to practical limitations of these methodologies, however, such as cost, manpower requirements, instrumentation footprint, and turnaround-time, MLVA provides the next most effective format using a 24- or 25-target multiplexed PCR system [2, 78]. Other PCR-based assays have also provided informative results, and in combination with work described here, PCR methods, in particular RD1 [90], RD_{Spain} (Dempsey *et al.*, #1 in preparation), and $CR_{holarctica}10$ (Dempsey *et al.*, #2 in preparation), demonstrate great resolving power. While the significance of this work is in its applicability to medical, forensic, micro-evolutionary science, the main application of genotyping and differentiating *F. tularensis* is to expediently identify the subspecies and subsequent subtypes to ensure adequate precautions are taken to minimize morbidity and mortality. Since phylogeographic variation has been described for *F. tularensis*, differentiation at that level provides forensically important intelligence, such as the ability to identify geographical source-strain populations of a potential naturally occurring outbreak or covert attack. In considering a strategy for expedient testing it is important to consider the types of specimens submitted as well as the documentation provided, such as those from clinical sources and those from the environment; and also to consider regulations imposed on testing each type. For example, clinical samples must be processed in accordance with Federal law, i.e., under College of American Pathologists (CAP) standards as well as by CDC-LRN guidelines, whereas testing of environmental or research samples doesn't have such stringent standards. Within the DoD, the Air Force, for example, has adopted uniform protocols and standards for testing as required by its Biological Augmentation Teams (BATs) and Homeland Defense Laboratory Response Teams (HLD-LRTs). To expediently satisfy the requirements of testing both clinical and environmental samples, a new PCR-based *F. tularensis* genotype differentiation algorithm is proposed.

5.6 Expeditionary *F. tularensis* PCR-Based Genotype Differentiation Algorithm:

The term "expeditionary" relates to an expedition, defined by Merriam-Webster Online (@ http://www.m-w.com/dictionary/expedition) as an excursion undertaken for a specific mission, and implies speed, or expediency, as a requirement. Such has been the mission of our DoD Expeditionary Medical Services (EMEDS) as well as their civilian Emergency Medical Services (EMS) First Responders counterparts, and therefore, expediency must be incorporated into their respective laboratory testing strategies, such as in identifying and genetically differentiating *F. tularensis* and other Category-A select agents, in order to minimize morbidity and mortality.

In the proposed model (see figure 5-3), regardless of sample type, an attempt should first be made to recover the isolate(s), which for *F. tularensis* is often challenging. Current DoD/AF EMEDS have this capability. Recovery enhancement can be improved by adapting a method recently demonstrated by Petersen *et al.* to enhance *F. tularensis* recovery by immediate plating onto cysteine heart agar with chocolatized 9% sheep blood (CHAB) supplemented with 7.5 mg of collistin, 2.5 mg of amphotericin, 0.5 mg of lincomycin, 4 mg of trimethoprim, and 10 mg of ampicllin per liter (CHAB-A) [114]. According to their study, the antibiotics serve to preserve the viability of *F. tularensis* as well as suppress growth of inhibitory bacterial species. Also shown in the study, freezing of tissues and expeditious transport to the laboratory seemed to help enhance recovery.

When a clinical sample is submitted, islolate-recovery enhancement will be performed as described above, and the sample will be processed using CDC-LRN Level-A or HLD-LRT protocols to provide a species-level identification, but which may still require several days to recover isolates in culture. Once a species-level identification is made, the next step will be to provide a subspecies identification using RD1 PCR. If the subspecies identified is subsp. *novicida* (likely from the U.S., and rarely from Australia), subsp. *mediaasiatica* (likely from Central Asia), or subsp. *holarctica-japonica* (from Japan), since no additional subspecies-subtype

PCR assays are available, MLVA-PCR will be performed to differentiate the strain-level genetic subtype. Note that due to the large size of necessary equipment for MLVA-PCR, MLVA may not be possible within EMEDS-level laboratories and may need to be referred to a local or theater-level reference laboratory. In the event the strain identified is from the *tularensis* subspecies, further testing using $CR_{holarctica}10$ PCR will provide subtyping as A1 correlating with A.I. (primarily from central and Eastern United States, but also from California, British Columbia, and Alaska) and A2 correlating with A.II. (primarily from Western States, Ontario, Texas, and per PESM analysis, also from Alaska). $CR_{holarctica}10$ PCR will also identify the newly described taxon, subsp. *neotularensis* (thus far only from Wyoming). Likewise, in the event RD1 identifies the strain as subsp. *holarctica*, RD_{Spain}-PCR will be performed to differentiate between strains that are wild-type *holarctica* (ubiquitous in the N. Hemisphere, and with minor virulence in humans) and those that are *holarctica*-RD_{Spain} (currently apparently restricted to the Iberian Peninsula in Spain and France, and potentially more virulent in humans than wild-type *holarctica* strains). MLVA will also be employed for further differentiating the strain-level genetic subtypes of subsps. *holarctica* and *tularensis* strains.

If the sample submitted is from an environmental source, certain steps in the differentiation algorithm may be modified or bypassed (which may also occasionally occur for clinical samples). If *F. tularensis* is suspected (as for a suspected biological release or outbreak), the need for culture confirmation may be delayed or circumvented provided DNA is available. In such a case, in the event that isolates are not culturable, attempts will be made to extract or whole-genome amplify DNA from the sample for RD1 PCR (subspecies-specific) and subsequent $CR_{holarctica}10$, RD_{Spain} (subspecies-subtype differential) and MLVA (subspecies and strain-subtype differential) PCR assays. If DNA cannot be measurably recovered from the sample, direct amplification using PCR for 16S rDNA, *F. tularensis* species-specific primers (i.e., *fop*A/*tul*4), RD1, $CR_{holarctica}10$, and RD_{Spain} may yield an identification as was discussed in chapter 1 for 16S

rDNA PCR. MLVA will also be performed since it is PCR-based and may amplify trace amounts of DNA specific to the MLVA targets.

5.7 Future Studies and Applicability to Other Organisms:

5.7.1 Whole Genome Sequencing (WGS):

Following preparation of DNA from the French *F.tularensis*-bacteremia strain [101] found to be RD_{Spain}-Positive in chapter 3, an aliquot of that DNA was submitted for whole genome sequencing through Los Alamos National Laboratories (LANL). Since that time, an approximately 15X coverage draft sequence has been completed, consisting of 5 contigs; and further sequencing to provide total gap closure has been approved for the project. I have either begun or have planned future collaborations involving this sequence.

In addition to the French *F. tularensis* isolate DNA, we will be submitting DNA from other organisms, including additional *F. tularensis* strains, for sequencing. Included among the *F. tularensis* strains is an isolate from a Nebraska human clinical tularemia case. Since that strain is a subsp. *tularensis*, it should be interesting to see how it compares with the published SCHU S4 genome sequence.

Based on pending completion of the French *F. tularensis* sequence, collaboration has been initiated with the Laboratory at Northern Arizona University under the direction of Dr. Paul Keim for the purpose of data-mining for single nucleotide polymorphisms (SNPs). SNPs (reviewed in REF.[115]) are single base pair positions in genomic DNA at which different sequence alternatives (alleles) exist in normal strains in some population(s). The low rate ($\sim 10^{-8}$ changes per nucleotide per generation) and essentially random nature of base changing events make such single-base alleles very evolutionarily stable and unlikely to mutate again to either a novel or ancestral state. Their rarity makes SNPs very important diagnostic markers, for example, in *Bacillus anthracis* [108], and suggest unique origins, as it is likely that each point

mutation occurred only once in the phylogenetic history of the species. With respect to VNTRs, their increased level of genetic diversity compared with SNPs is not only a function of differences in having faster mutation rates, but also in the number of possible allelic states due to involvement of more nucleotides in the repeat unit as well as variability in the number of repeats [108]. Our goal of SNP discovery in the French isolate WGS, and subsequently in the Nebraska isolate WGS, is to include the data in the database for comparative SNP positions in other *F. tularensis* strains sequenced, which should then facilitate designing *F. tularensis*-specific SNP-based differential RT-PCR and microarray assays.

In addition to the collaboration at NAU, future work is planned with a colleague at The Institute for Genomic Research (TIGR) to completely annotate the French *F. tularensis* sequence once it is closed. This project holds promise to be the first completely annotated subsp. *holarctica* sequence published. The sequence will be scaffolded and compared with the published SCHU S4 sequence, and should provide a much more detailed subsps. *holarctica* : *tularensis* genome comparison than our PESM study. Also, we should be able to identify other features at the whole-genome level which would support the strain's observed phenotypic characteristics involving its interesting epidemiology and known RD_{Spain} polymorphism.

5.7.2 Non-WGS Projects:

Apart from WGS, several other molecular methodologies may be applied as previously discussed throughout this dissertation, and each has its own unique advantages and disadvantages depending on the organism evaluated and users' research objectives. For example, many organisms have been found to contain mobile genetic elements which have been shown to facilitate rearrangement and even horizontal transfer, i.e., bacteria phages, integrons, transposons/IS elements as described here at length for *Francisella tularensis* [91, 116]. Also included here are plasmids/plasmid-associated genes, such as those found in the *Bacillus cereus* group which have been known to horizontally transfer and even cause high morbidity [117] and

death [118] in typically non-pathogenic strains. In such cases, methodologies to evaluate basic core genomic content are not sufficient, thus necessitating the analysis of accessory genomic content, i.e., plasmids, as well as genomic features such as pathogenicity islands, IS element-mediated rearrangements including insertions/deletions, etc. In the case of genomes such as *F. tularensis*, such rearrangements may alter pathogenesis and/or diminish utility of current molecular methods, such as PCR, due to the potential translocation of one or the other primer disallowing formation of an amplicon. In spite of that potential, it is noteworthy that our interesting RD_{Spain} and Wyoming CR10-A3/subsp. '*neotularensis*' strains produced *F. tularensis* wild-type PCR results with our biodefense assays (see figure 4-10). To eliminate this potential dilemma, the maximum number of strains should be evaluated with any given method to observe all the possible genotypes; and such a strain collection should include as many strains from spatially and geographically diverse genetic backgrounds as has been attempted for our *F. tularensis* strain collection for the work in chapters 3 and 4.

Increasing the amount of genome coverage for analysis improves resolution, but this can also be achieved using multiple targets or by combining methodologies as has been described in chapter 1. This approach, when applied to molecular methods, has recently been termed as "Progressive hierarchical resolving assays using nucleic acids" (PHRANA)[108]. Molecular methodologies other than nucleic acid analysis should also prove beneficial for overall strain characterization, and therefore are described below. Ultimately, proper combinations of targets and/or methodologies should allow maximum identification and differentiation of each strain within a given collection.

5.7.3 Collaborations and applicability to other organisms:

With PHRANA in mind, a comparison of molecular methods is currently being performed at the AFIP in collaboration with the University of Nebraska Medical Center for

characterizing a diverse strain collection of twenty *F. tularensis* isolates. The molecular methods being performed at the AFIP include AFLP, MLVA using eight of the published targets for *F. tularensis* (otherwise called "mini-MLVA"), and an enhanced REP-PCR method which has recently been used successfully for other organisms [119]. Fatty acid profiling using the MIDI Sherlock™ system, which is not a genotyping method but has been shown to differentiate among highly conserved organisms such as the *B. cereus* group of organisms, is also being employed for the current *F. tularensis* study. The MIDI system is a useful tool for monitoring possible genetic changes in an organism during laboratory manipulations.

Currently, the AFIP and UNMC laboratories are collaboratively studying differential proteomics of *Yersinia pestis*, and in particular, how that species differs from other *Yersinia* species at the proteome level. The study has shown excellent promise of allowing differentiation at the species level based on their unique Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectromotry (MALDI-TOF-MS) profiles. In addition, we are very close to identifying species-specific proteins by Surface-Enhanced Time-of-Flight Mass Spectrometry (SELDI-TOF-MS). With the *Y. pestis* proteomics study as a model system, we should be able to duplicate success for *F. tularensis* as well, and soon it will be possible to correlate all of the different findings into a comprehensive understanding of *F. tularensis* biology and genetics.

The CGH microarray and PESM models for *F. tularensis*, as well as identification of subsequent molecular targets and design of their respective PCR assays presented here should provide excellent applicability to other organisms. For example, collaborative CGH studies are on-going between the AFIP and UNL laboratories, and have thus far provided several species-specific genetic targets for designing new differential PCR assays. By applying the resources of these methods with the other PHRANA methods previously described and those yet to be introduced, we should be able to provide comprehensive molecular characterization and differentiation for numerous bacterial organisms, as well as identify genomic differences

potentially of biological importance. This comprehensive strategy should prove useful as DoD and other government agencies are gearing up to literally study the pathogens of the world at the WGS and other molecular-based levels.

The End!

Figure 5-1.

Model of Molecular-Basis of Divergence of *F. tularensis*.

The model shows divergence of the respective subspecies of *F. tularensis* from the representative *F. tularensis* ancestral clone. Light-blue boxes represent mechanisms of divergence previously described in the literature whereas orange boxes represent *de novo* mechanisms of divergence described in this dissertation.

Figure 5-1: Model of Molecular-Basis of Divergence of *F. tularensis*.

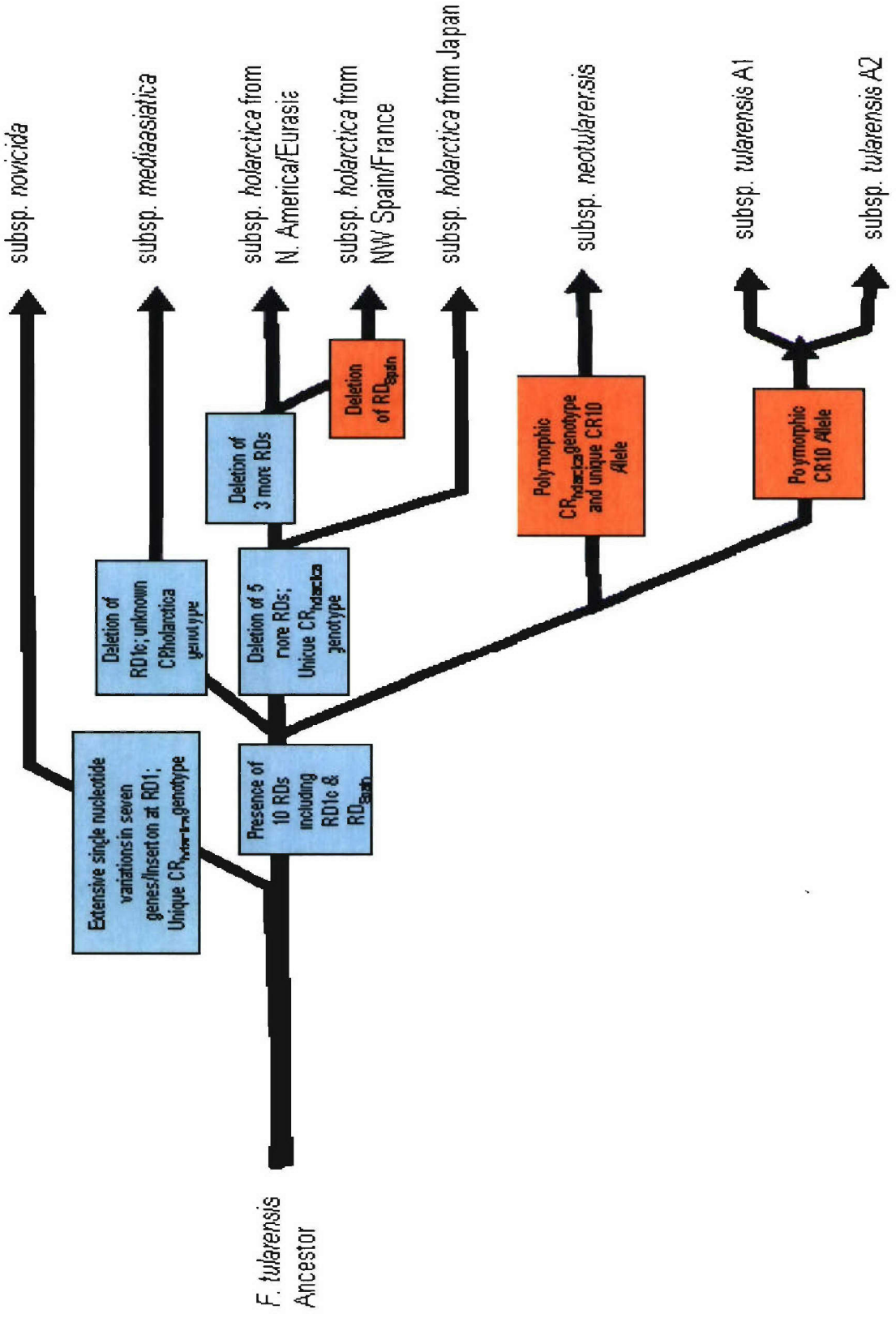


Figure 5-2.

Current Global Divergence of *F. tularensis.*

The figure shows the currently known locations of the various subspecies of *F. tularensis* (legend shown at the base of the map. Notice the single location of the *de novo* subspecies, *neotularensis*, located in Wyoming, in a geographic region also comprised of subsp. *holarctica*, *novicida*, and *tularensis*-A2.

Figure 5-2: Current Global Divergence of *F. tularensis*

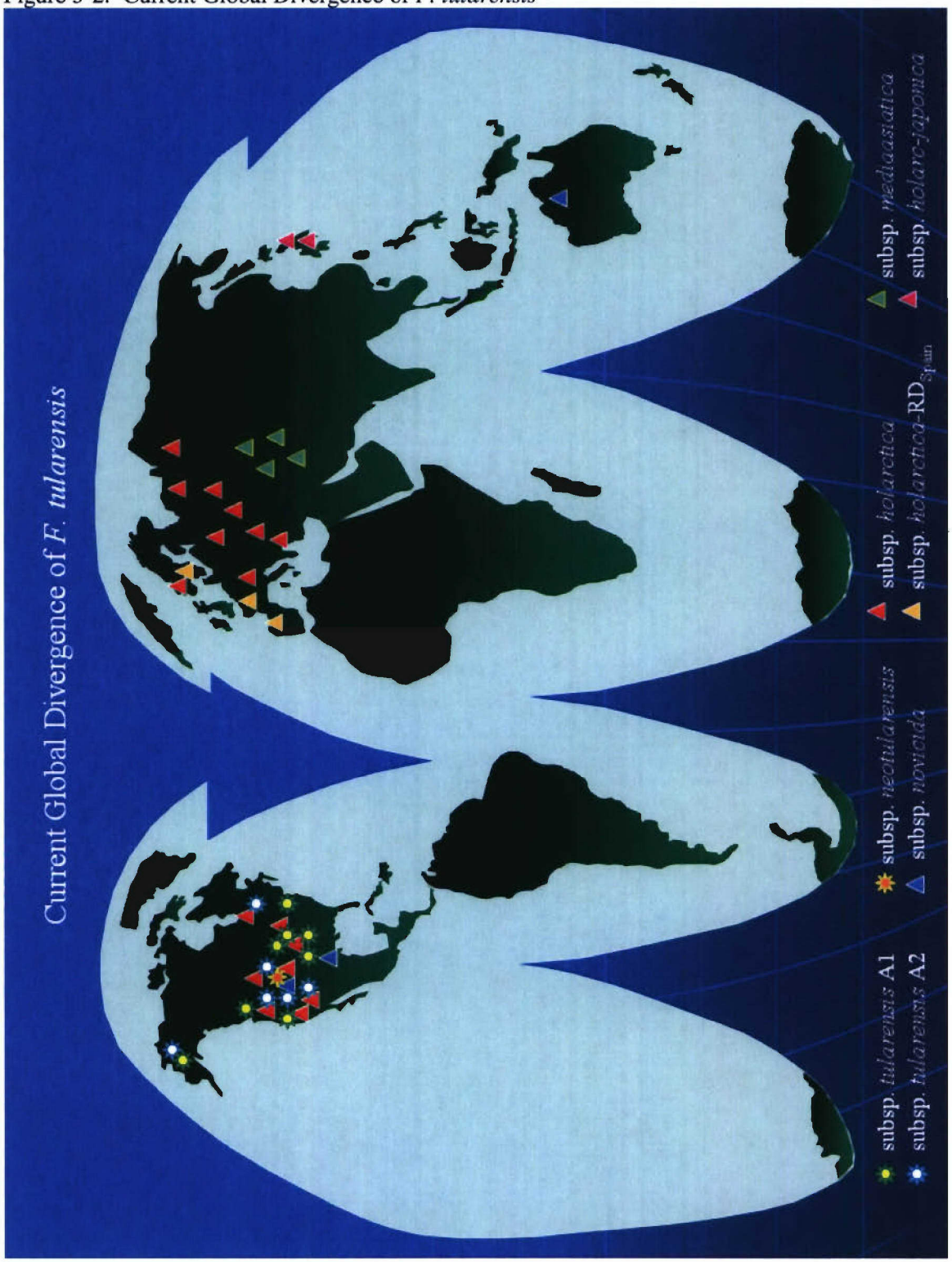
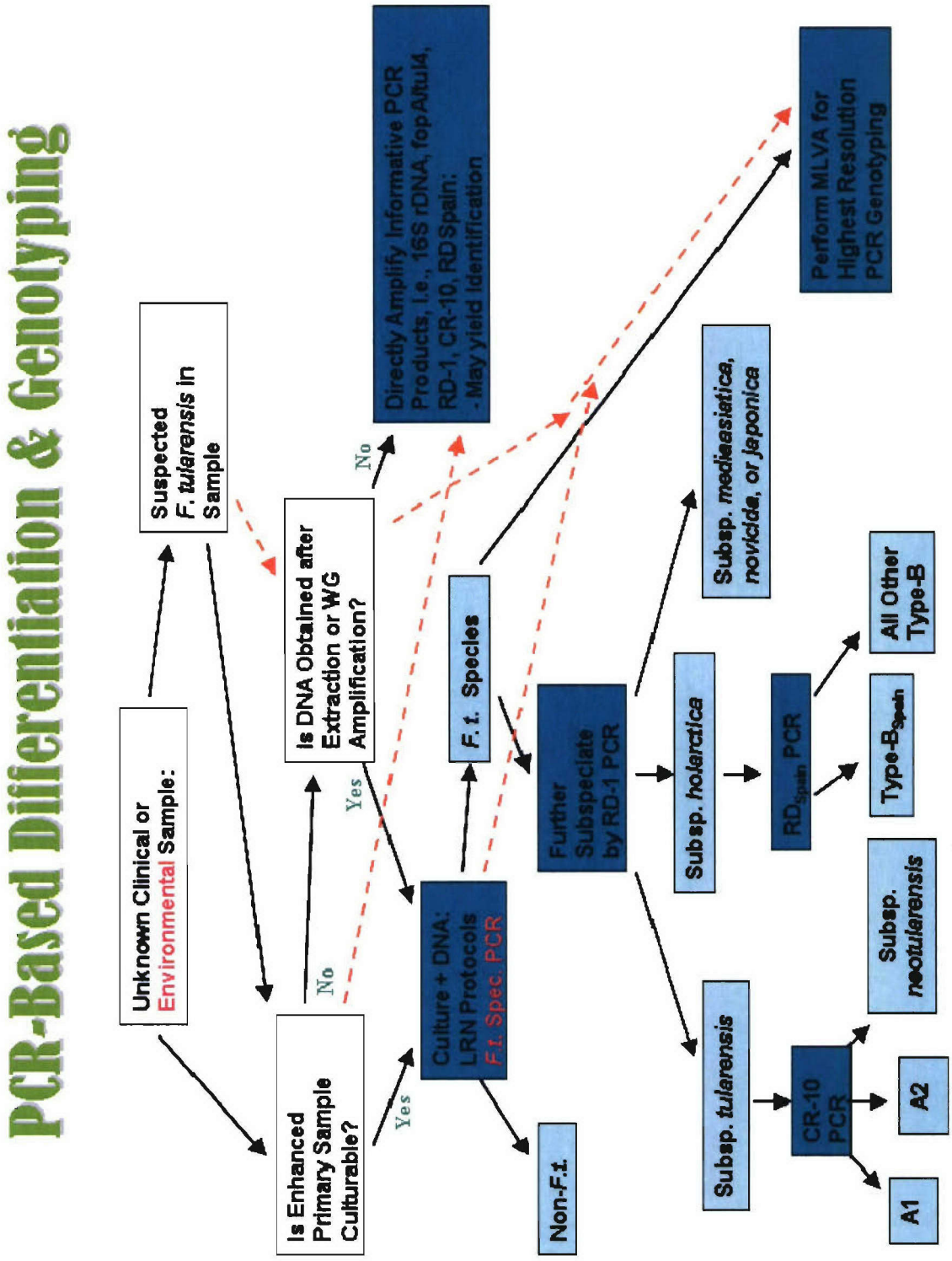


Figure 5-3.

Expeditionary *F. tularensis* PCR-based Genotype Differentiation Algorithm.

White boxes in the figure show pre-analytical/decision-making steps, aquamarine boxes show laboratory procedural steps, and light-blue boxes show testing outcomes. The orange-dashed arrowed lines indicate alternative steps for processing environmental (but in some cases, clinical) samples.

Figure 5-3: Expeditionary *F. tularensis* PCR-based Genotype Differentiation Algorithm.



References:

1. McCoy, G.W. and C.W. Chapin., *Further observations on a plague-like disease of rodents with a preliminary note on the causative agent, Bacterium tularense.* J. Infect. Dis, 1912. **10**: p. 61-72.

2. Johansson, A., et al., *Worldwide genetic relationships among Francisella tularensis isolates determined by multiple-locus variable-number tandem repeat analysis.* J Bacteriol, 2004. **186**(17): p. 5808-18.

3. Hopla, C.E. and A.K. Hopla, *Tularemia.* 2nd ed. Handbook of Zoonoses, ed. G.W. Beran. 1994, Boca Raton, FL: CRC Press. 113-126.

4. Evans, M.E. and A.M. Friedlander, *Chapter 24: Tularemia*, in *Text Book of Military Medicine Part I: Medical Aspects of Chemical and Biological Warfare*, R. Zajtchuk and R. Belamy, Editors. 1997, Office of the Surgeon General: Washington, D.C. p. 10.

5. Alibek, K., *Biohazard.* 1999, New York: Random House.

6. Harris, S., *Japanese biological warfare research on humans: a case study of microbiology and ethics.* Ann N Y Acad Sci, 1992. **666**: p. 21-52.

7. Dennis, D.T., et al., *Tularemia as a biological weapon: medical and public health management.* Jama, 2001. **285**(21): p. 2763-73.

8. CDC-Website-http://www.bt.cdc.gov/agent/agentlist-category.asp, *Categories of Bioterrorism Agents/Diseases*. 2004, Centers for Disease Control and Prevention: Atlanta, GA.

9. Jernigan, D.B., et al., *Investigation of bioterrorism-related anthrax, United States, 2001: epidemiologic findings.* Emerg Infect Dis, 2002. **8**(10): p. 1019-28.

10. Oyston, P.C., A. Sjostedt, and R.W. Titball, *Tularaemia: bioterrorism defence renews interest in Francisella tularensis.* Nat Rev Microbiol, 2004. **2**(12): p. 967-78.

11. Organization, W.H., *Health Aspects of Chemical and Biological Weapons*. 1970, Geneva: WHO.

12. (CDC), C.f.D.C.a.P., *Basic Protocols for Level A Laboratories for the Presumptive Identification of Francisella tularensis* 2001: Atlanta, GA.

13. Chu, M.C. and R.S. Weyant, *Francisella and Brucella*, in *Manual of Clinical Microbiology, 8th Edition*, P.R. Murray, et al., Editors. 2003, ASM Press: Washington, D.C. p. 20.

14. Svensson, K., et al., *Evolution of subspecies of Francisella tularensis.* J Bacteriol, 2005. **187**(11): p. 3903-8.

15. Titball, R.W., A. Johansson, and M. Forsman, *Will the enigma of Francisella tularensis virulence soon be solved?* Trends Microbiol, 2003. **11**(3): p. 118-23.

16. USDHHS, *Biosafety in Microbiological and Biomedical Laboratories (BMBL) Biosafety in Microbiological and Biomedical Laboratories (BMBL) 4th Edition.* 1999, U.S. Government Printing Office: Washington, DC.

17. Petersen, J.M. and M.E. Schriefer, *Tularemia: emergence/re-emergence.* Vet Res, 2005. **36**(3): p. 455-467.

18. Jellison, W.L., *Tularemia in North America*, University of Montana, Missoula, Montana. 276 p.

19. Morner, T., *The ecology of tularaemia.* Rev Sci Tech, 1992. **11**(4): p. 1123-30.

20. Eliasson, H., et al., *The 2000 tularemia outbreak: a case-control study of risk factors in disease-endemic and emergent areas, Sweden.* Emerg Infect Dis, 2002. **8**(9): p. 956-60.

21. Abd, H., et al., *Survival and growth of Francisella tularensis in Acanthamoeba castellanii.* Appl Environ Microbiol, 2003. **69**(1): p. 600-6.

22. Sjöstedt, A., *Family XVII, Francisellaceae. Genus I, Franciesella*, in *Bergey's Manual of Systematic Bacteriology*, G.M. Garrity, Editor. 2003, Springer-Verlag: New York, N.Y. p. 111-113.

23. Whipp, M.J., et al., *Characterization of a novicida-like subspecies of Francisella tularensis isolated in Australia.* J Med Microbiol, 2003. **52**(Pt 9): p. 839-42.

24. Gurycova, D., *First isolation of Francisella tularensis subsp. tularensis in Europe.* Eur J Epidemiol, 1998. **14**(8): p. 797-802.

25. Sandstrom, G., et al., *Characterization and classification of strains of Francisella tularensis isolated in the central Asian focus of the Soviet Union and in Japan.* J Clin Microbiol, 1992. **30**(1): p. 172-5.

26. Thomas, R., et al., *Discrimination of human pathogenic subspecies of Francisella tularensis by using restriction fragment length polymorphism.* J Clin Microbiol, 2003. **41**(1): p. 50-7.

27. Olsufjev, N.G. and I.S. Meshcheryakova, *Infraspecific taxonomy of tularemia agent Francisella tularensis McCoy et Chapin.* J Hyg Epidemiol Microbiol Immunol, 1982. **26**(3): p. 291-9.

28. Perez-Castrillon, J.L., et al., *Tularemia epidemic in northwestern Spain: clinical description and therapeutic response.* Clin Infect Dis, 2001. **33**(4): p. 573-6.

29. Bachiller Luque, P., et al., *[Preliminary report of an epidemic tularemia outbreak in Valladolid].* Rev Clin Esp, 1998. **198**(12): p. 789-93.

30. Bellido-Casado, J., et al., *Report on five cases of tularaemic pneumonia in a tularaemia outbreak in Spain.* Eur J Clin Microbiol Infect Dis, 2000. **19**(3): p. 218-20.

31. García Peña, F.J., et al., *An outbreak of tularemia in Castilla-Leon. First isolation of Francisella tularensis in Spain.* Medicina Veterinaria, 1998. **15**: p. 418-423.

32. de la Puente-Redondo, V.A., et al., *Comparison of different PCR approaches for typing of Francisella tularensis strains.* J Clin Microbiol, 2000. **38**(3): p. 1016-22.

33. Garcia Del Blanco, N., et al., *Genotyping of Francisella tularensis strains by pulsed-field gel electrophoresis, amplified fragment length polymorphism fingerprinting, and 16S rRNA gene sequencing.* J Clin Microbiol, 2002. **40**(8): p. 2964-72.

34. Garcia del Blanco, N., et al., *Biochemical characterisation of Francisella tularensis strains isolated in Spain.* Vet Rec, 2004. **154**(2): p. 55-6.

35. Anda, P., et al., *Waterborne outbreak of tularemia associated with crayfish fishing.* Emerg Infect Dis, 2001. **7**(3 Suppl): p. 575-82.

36. Evans, M.E., et al., *Tularemia: a 30-year experience with 88 cases.* Medicine (Baltimore), 1985. **64**(4): p. 251-69.

37. Fortier, A.H., et al., *Activation of macrophages for destruction of Francisella tularensis: identification of cytokines, effector cells, and effector molecules.* Infect Immun, 1992. **60**(3): p. 817-25.

38. Elkins, K.L., et al., *Rapid generation of specific protective immunity to Francisella tularensis.* Infect Immun, 1992. **60**(11): p. 4571-7.

39. Elkins, K.L., et al., *T-cell-independent resistance to infection and generation of immunity to Francisella tularensis.* Infect Immun, 1993. **61**(3): p. 823-9.

40. Lai, X.H., I. Golovliov, and A. Sjostedt, *Francisella tularensis induces cytopathogenicity and apoptosis in murine macrophages via a mechanism that requires intracellular bacterial multiplication.* Infect Immun, 2001. **69**(7): p. 4691-4.

41. Anthony, L.S., et al., *The requirement for gamma interferon in resistance of mice to experimental tularemia.* Microb Pathog, 1989. **7**(6): p. 421-8.

42. Leiby, D.A., et al., *In vivo modulation of the murine immune response to Francisella tularensis LVS by administration of anticytokine antibodies.* Infect Immun, 1992. **60**(1): p. 84-9.

43. Clemens, D.L., B.Y. Lee, and M.A. Horwitz, *Virulent and avirulent strains of Francisella tularensis prevent acidification and maturation of their phagosomes and escape into the cytoplasm in human macrophages.* Infect Immun, 2004. **72**(6): p. 3204-17.

44. Janeway, C., et al., *Immunobiology.* 5th ed. The Immune System in Health and Disease. ed. P. Austin, et al. 2001, New York, NY: Garland Publishing. 732 pages.

45. Forestal, C.A., et al., *Francisella tularensis selectively induces proinflammatory changes in endothelial cells.* J Immunol, 2003. **171**(5): p. 2563-70.

46. Fortier, A.H., et al., *Life and death of an intracellular pathogen: Francisella tularensis and the macrophage.* Immunol Ser, 1994. **60**: p. 349-61.

47. Fortier, A.H., et al., *Growth of Francisella tularensis LVS in macrophages: the acidic intracellular compartment provides essential iron required for growth.* Infect Immun, 1995. **63**(4): p. 1478-83.

48. Larsson, P., et al., *The complete genome sequence of Francisella tularensis, the causative agent of tularemia.* Nat Genet, 2005. **37**(2): p. 153-9.

49. Golovliov, I., et al., *An attenuated strain of the facultative intracellular bacterium Francisella tularensis can escape the phagosome of monocytic cells.* Infect Immun, 2003. **71**(10): p. 5940-50.

50. Lindgren, H., et al., *Factors affecting the escape of Francisella tularensis from the phagolysosome.* J Med Microbiol, 2004. **53**(Pt 10): p. 953-8.

51. Lai, X.H. and A. Sjostedt, *Delineation of the molecular mechanisms of Francisella tularensis-induced apoptosis in murine macrophages.* Infect Immun, 2003. **71**(8): p. 4642-6.

52. Maier, T.M., et al., *Construction and characterization of a highly efficient Francisella shuttle plasmid.* Appl Environ Microbiol, 2004. **70**(12): p. 7511-9.

53. Golovliov, I., et al., *A method for allelic replacement in Francisella tularensis.* FEMS Microbiol Lett, 2003. **222**(2): p. 273-80.

54. Golovliov, I., et al., *Identification of proteins of Francisella tularensis induced during growth in macrophages and cloning of the gene encoding a prominently induced 23-kilodalton protein.* Infect Immun, 1997. **65**(6): p. 2183-9.

55. Nano, F.E., et al., *A Francisella tularensis pathogenicity island required for intramacrophage growth.* J Bacteriol, 2004. **186**(19): p. 6430-6.

56. Gray, C.G., et al., *The identification of five genetic loci of Francisella novicida associated with intracellular growth.* FEMS Microbiol Lett, 2002. **215**(1): p. 53-6.

57. Telepnev, M., et al., *Francisella tularensis inhibits Toll-like receptor-mediated activation of intracellular signalling and secretion of TNF-alpha and IL-1 from murine macrophages.* Cell Microbiol, 2003. **5**(1): p. 41-51.

58. Lauriano, C.M., et al., *MglA regulates transcription of virulence factors necessary for Francisella tularensis intraamoebae and intramacrophage survival.* Proc Natl Acad Sci U S A, 2004. **101**(12): p. 4246-9.

59. Baron, G.S. and F.E. Nano, *MglA and MglB are required for the intramacrophage growth of Francisella novicida.* Mol Microbiol, 1998. **29**(1): p. 247-59.

60. Reilly, T.J., et al., *Characterization and sequencing of a respiratory burst-inhibiting acid phosphatase from Francisella tularensis.* J Biol Chem, 1996. **271**(18): p. 10973-83.

61. Anthony, L.S., et al., *Isolation of a Francisella tularensis mutant that is sensitive to serum and oxidative killing and is avirulent in mice: correlation with the loss of MinD homologue expression.* FEMS Microbiol Lett, 1994. **124**(2): p. 157-65.

62. McDonald, M.K., S.C. Cowley, and F.E. Nano, *Temperature-sensitive lesions in the Francisella novicida valA gene cloned into an Escherichia coli msbA lpxK mutant affecting deoxycholate resistance and lipopolysaccharide assembly at the restrictive temperature.* J Bacteriol, 1997. **179**(24): p. 7638-43.

63. Mdluli, K.E., et al., *Serum-sensitive mutation of Francisella novicida: association with an ABC transporter gene.* Microbiology, 1994. **140 (Pt 12)**: p. 3309-18.

64. Sandstrom, G., et al., *Immunogenicity and toxicity of lipopolysaccharide from Francisella tularensis LVS.* FEMS Microbiol Immunol, 1992. **5**(4): p. 201-10.

65. Craig, L., M.E. Pique, and J.A. Tainer, *Type IV pilus structure and bacterial pathogenicity.* Nat Rev Microbiol, 2004. **2**(5): p. 363-78.

66. Friedlander, A.M., *Chapter 22: Anthrax*, in *Text Book of Military Medicine Part I: Medical Aspects of Chemical and Biological Warfare*, C.M.Q. Lorraine B. Davis, Editor. 1997, Office of the Surgeon General: Washington, D.C. p. 11.

67. Ivins, B.E., et al., *Immunization studies with attenuated strains of Bacillus anthracis.* Infect Immun, 1986. **52**(2): p. 454-8.

68. Weisburg, W.G., et al., *16S ribosomal DNA amplification for phylogenetic study.* J Bacteriol, 1991. **173**(2): p. 697-703.

69. Schwartz, D.C. and C.R. Cantor, *Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis.* Cell, 1984. **37**(1): p. 67-75.

70. Vos, P., et al., *AFLP: a new technique for DNA fingerprinting.* Nucleic Acids Res, 1995. **23**(21): p. 4407-14.

71. Wyman, A.R. and R. White, *A highly polymorphic locus in human DNA.* Proc Natl Acad Sci U S A, 1980. **77**(11): p. 6754-8.

72. Olive, D.M. and P. Bean, *Principles and applications of methods for DNA-based typing of microbial organisms.* J Clin Microbiol, 1999. **37**(6): p. 1661-9.

73. van Belkum, A., *The role of short sequence repeats in epidemiologic typing.* Curr Opin Microbiol, 1999. **2**(3): p. 306-11.

74. Keim, P., et al., *Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within Bacillus anthracis.* J Bacteriol, 2000. **182**(10): p. 2928-36.

75. Klevytska, A.M., et al., *Identification and characterization of variable-number tandem repeats in the Yersinia pestis genome.* J Clin Microbiol, 2001. **39**(9): p. 3179-85.

76. Farlow, J., et al., *Francisella tularensis strain typing using multiple-locus, variable-number tandem repeat analysis.* J Clin Microbiol, 2001. **39**(9): p. 3186-92.

77. Johansson, A., et al., *Extensive allelic variation among Francisella tularensis strains in a short-sequence tandem repeat region.* J Clin Microbiol, 2001. **39**(9): p. 3140-6.

78. Farlow, J., et al., *Francisella tularensis in the United States.* Emerg Infect Dis, 2005. **11**(12): p. 1835-1841.

79. Forsman, M., G. Sandstrom, and A. Sjostedt, *Analysis of 16S ribosomal DNA sequences of Francisella strains and utilization for determination of the phylogeny of the genus and for identification of strains by PCR.* Int J Syst Bacteriol, 1994. **44**(1): p. 38-46.

80. Forsman, M., G. Sandstrom, and B. Jaurin, *Identification of Francisella species and discrimination of type A and type B strains of F. tularensis by 16S rRNA analysis.* Appl Environ Microbiol, 1990. **56**(4): p. 949-55.

81. Barns, S.M., et al., *Detection of diverse new Francisella-like bacteria in environmental samples.* Appl Environ Microbiol, 2005. **71**(9): p. 5494-500.

82. Desai, M., et al., *Fluorescent amplified-fragment length polymorphism analysis of an outbreak of group A streptococcal invasive disease.* J Clin Microbiol, 1998. **36**(11): p. 3133-7.

83. Sreevatsan, S., et al., *Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination.* Proc Natl Acad Sci U S A, 1997. **94**(18): p. 9869-74.

84. Stanley, J. and N. Saunders, *DNA insertion sequences and the molecular epidemiology of Salmonella and Mycobacterium.* J Med Microbiol, 1996. **45**(4): p. 236-51.

85. Achtman, M., et al., *Yersinia pestis, the cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis.* Proc Natl Acad Sci U S A, 1999. **96**(24): p. 14043-8.

86. Kremer, K., et al., *Comparison of methods based on different molecular epidemiological markers for typing of Mycobacterium tuberculosis complex strains: interlaboratory study of discriminatory power and reproducibility.* J Clin Microbiol, 1999. **37**(8): p. 2607-18.

87. Fulop, M., D. Leslie, and R. Titball, *A rapid, highly sensitive method for the detection of Francisella tularensis in clinical samples using the polymerase chain reaction.* Am J Trop Med Hyg, 1996. **54**(4): p. 364-6.

88. Long, G.W., et al., *Detection of Francisella tularensis in blood by polymerase chain reaction.* J Clin Microbiol, 1993. **31**(1): p. 152-4.

89. Sjostedt, A., et al., *Detection of Francisella tularensis in ulcers of patients with tularemia by PCR.* J Clin Microbiol, 1997. **35**(5): p. 1045-8.

90. Broekhuijsen, M., et al., *Genome-wide DNA microarray analysis of Francisella tularensis strains demonstrates extensive genetic conservation within the species but identifies regions that are unique to the highly virulent F. tularensis subsp. tularensis.* J Clin Microbiol, 2003. **41**(7): p. 2924-31.

91. Samrakandi, M.M., et al., *Genome diversity among regional populations of Francisella tularensis subspecies tularensis and Francisella tularensis subspecies holarctica isolated from the US.* FEMS Microbiol Lett, 2004. **237**(1): p. 9-17.

92. (CDC), C.f.D.C.a.P. *Where in the World is Tularemia?* in *Tularemia Update: Epidemiology and Ecology of the Engimatic Zoonotic Agent* 2004. Denver, CO.

93. Wilson, K., *Preparation of Genomic DNA from Bacteria*, in *Current Protocols in Molecular Biology*, F.M. Ausubel, et al., Editors. 1997, John Wiley & Sons. p. 2.4.1-2.4.5.

94. Zhang, C., et al., *Genome diversification in phylogenetic lineages I and II of Listeria monocytogenes: identification of segments unique to lineage II populations.* J Bacteriol, 2003. **185**(18): p. 5573-84.

95. Li, W.-H., *Molecular Evolution.*, Sinauer, Sunderland, and MA.. Editors. 1997. p. 106-108.

96. Kim, J., et al., *Ancestral divergence, genome diversification, and phylogeographic variation in subpopulations of sorbitol-negative, beta-glucuronidase-negative enterohemorrhagic Escherichia coli O157.* J Bacteriol, 2001. **183**(23): p. 6885-97.

97. Swofford, D.L., *PAUP*, Phylogenetic Analysis Using Parsimony (and Other Methods). Version 4.* 2001, Sunderland, Massachusetts: Sinauer Associates.

98. Lech, K. and R. Brent, *Plating Lambda Phage to Generate Plaques*, in *Current Protocols in Molecular Biology*, F.M. Ausubel, et al., Editors. 1990: USA.

99. Lech, K. and R. Brent, *Media Preparation and Bacteriological Tools*, in *Current Protocols in Molecular Biology*, F.M. Ausubel, et al., Editors. 2002: USA.

100. Ellis, J., et al., *Tularemia.* Clin Microbiol Rev, 2002. **15**(4): p. 631-46.

101. Haristoy, X., et al., *Francisella tularensis bacteremia.* J Clin Microbiol, 2003. **41**(6): p. 2774-6.

102. Johansson, A., et al., *Evaluation of PCR-based methods for discrimination of Francisella species and subspecies and development of a specific PCR that distinguishes the two major subspecies of Francisella tularensis.* J Clin Microbiol, 2000. **38**(11): p. 4180-5.

103. Gutierrez, M.P., et al., *Serologic evidence of human infection by Francisella tularensis in the population of Castilla y Leon (Spain) prior to 1997.* FEMS Immunol Med Microbiol, 2003. **35**(2): p. 165-9.

104. Chain, P.S., et al., *Insights into the evolution of Yersinia pestis through whole-genome comparison with Yersinia pseudotuberculosis.* Proc Natl Acad Sci U S A, 2004. **101**(38): p. 13826-31.

105. Bumbaugh, A.C., et al. *Genomic variation in Shigella dysenteriae type I using a new approach called paired end sequencing mapping.* in *American Society for Microbiology 103rd General Meeting.* 2003. Washington, DC: ASM Press.

106. Cronan, J.E., Jr., *Interchangeable enzyme modules. Functional replacement of the essential linker of the biotinylated subunit of acetyl-CoA carboxylase with a linker from the lipoylated subunit of pyruvate dehydrogenase.* J Biol Chem, 2002. **277**(25): p. 22520-7.

107. Gil, H., J.L. Benach, and D.G. Thanassi, *Presence of pili on the surface of Francisella tularensis.* Infect Immun, 2004. **72**(5): p. 3042-7.

108. Keim, P., et al., *Anthrax molecular epidemiology and forensics: using the appropriate marker for different evolutionary scales.* Infect Genet Evol, 2004. **4**(3): p. 205-13.

109. Arbeit, R.D., et al., *Resolution of recent evolutionary divergence among Escherichia coli from related lineages: the application of pulsed field electrophoresis to molecular epidemiology.* J Infect Dis, 1990. **161**(2): p. 230-5.

110. Scott, L., et al., *The characterisation of E. coli O157:H7 isolates from cattle faeces and feedlot environment using PFGE.* Vet Microbiol, 2006.

111. Brosch, R., et al., *A new evolutionary scenario for the Mycobacterium tuberculosis complex.* Proc Natl Acad Sci U S A, 2002. **99**(6): p. 3684-9.

112. Mostowy, S., et al., *Genomic deletions suggest a phylogeny for the Mycobacterium tuberculosis complex.* J Infect Dis, 2002. **186**(1): p. 74-80.

113. Francis, E., *Sources of infection and seasonal incidence of tularemia in man*, in *Public Health Report 52*. 1937. p. 103-113.

114. Petersen, J.M., et al., *Methods for enhanced culture recovery of Francisella tularensis.* Appl Environ Microbiol, 2004. **70**(6): p. 3733-5.

115. Brookes, A.J., *The essence of SNPs.* Gene, 1999. **234**(2): p. 177-86.

116. Michael Chandler, J.M., *Chapter 15: Insertion Sequences Revisited*, in *Mobile DNA II*, e.a. Nancy L. Craig, Editor. 2002, ASM Press: Washington, D.C. p. 61.

117. Hoffmaster, A.R., et al., *Identification of anthrax toxin genes in a Bacillus cereus associated with an illness resembling inhalation anthrax.* Proc Natl Acad Sci U S A, 2004. **101**(22): p. 8449-54.

118. Miller, J.M., et al., *Fulminating bacteremia and pneumonia due to Bacillus cereus.* J Clin Microbiol, 1997. **35**(2): p. 504-7.

119. Jones, S.W., et al., *DNA assays for detection, identification, and individualization of select agent microorganisms.* Croat Med J, 2005. **46**(4): p. 522-9.